

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 19, 2002, 06:53:36 ; Search time 25 Seconds

(without alignments)
709.342 Million cell updates/sec

Title: US-09-807-459-2

Perfect score: 2359
Sequence: 1 MAPSDVGDVTKTLAASES.....DPKALIRKSTEADNLLK 458

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	5.4	607	YMH1_YEAST	Q04632 saccharomyc
2	122	5.2	1829	FRPC_NEIMC	P55127 neisseria m
3	121	5.1	566	HEMA_IAUSS	P03453 influenza a
4	120.5	5.1	566	HEMA_IAZUK	P26141 influenza a
5	117	5.0	1038	CIN8_YEAST	P27895 saccharomyc
6	117	5.0	1829	FRPC_NEIMB	Q919V5 neisseria m
7	116	4.9	437	Y162_UREPA	Q02455 saccharomyc
8	115	4.9	1875	MLP1_YEAST	P08799 dictyosteli
9	113.5	4.8	2116	MY52_DICDI	P15205 rattus norv
10	113.5	4.8	2459	HEMA_TALEN	P18876 influenza a
11	112	4.7	566	Y086_SCHPO	Q10411 schizosacch
12	112	4.7	1957	Y086_SCHPO	Q949Q2 schizosacch
13	111.5	4.7	832	ALP6_SCHPO	P54788 kluyveromyc
14	110.5	4.7	886	ORC1_KLUDA	P06808 b bacillari
15	110.5	4.7	6359	BACQ_BACLI	P03452 influenza a
16	110	4.7	566	HEMA_IAPUE	Q09684 schizosacch
17	110	4.7	577	THRI_SCHPO	P12587 influenza a
18	110	4.7	1271	Y338_MYCCE	P47580 mycoplasma
19	108	4.6	566	HEMA_IATLE	Q05547 methanococ
20	107.5	4.6	462	YB47_METUA	P11134 influenza a
21	107	4.5	550	HEMA_IATN3	Q06431 dugbe virus
22	106.5	4.5	4036	RRPL_DUGBY	P12587 influenza a
23	106	4.5	550	HEMA_IADH6	P12588 influenza a
24	106	4.5	550	HEMA_IADH7	P34074 saccharomyc
25	106	4.5	624	YMO8_YEAST	P43257 influenza a
26	105.5	4.5	550	HEMA_IADHK	P75556 mycoplasma
27	105.5	4.5	566	HEMA_IATN7	P03440 influenza a
28	105.5	4.5	1030	Y075_MYCPN	P75556 mycoplasma
29	105.5	4.5	2670	YAO5_SCHPO	Q10105 schizosacch
30	105	4.5	808	Y066_NPVAC	P41467 autographa
31	105	4.5	857	NFM_CHICK	P16053 gallus gall
32	104.5	4.4	778	YF05_METUA	Q05900 methanococ
33	104	4.4	550	HEMA_IADH1	P12582 influenza a

ALIGNMENTS

34	104	4.4	550	1	HEMA_IADH2	P12583 influenza a
35	104	4.4	550	1	HEMA_IADH3	P12584 influenza a
36	103.5	4.4	2464	1	MAPB_MOUSE	P14873 mus musculu
37	103	4.4	566	1	HEMA_IADH2	P26135 influenza a
38	103	4.4	1577	1	MYSH_ACACA	P47808 acanthamoeb
39	102.5	4.3	1485	1	TOP2_SCHPO	P08096 schizosacch
40	102.5	4.3	3106	1	IMA2_MOUSE	O60675 mus musculu
41	102	4.3	540	1	YK26_YEAST	P36112 saccharomyc
42	102	4.3	1664	1	INT1_CANAL	P53705 candida alb
43	101.5	4.3	566	1	HEMA_IAME1	P03449 influenza a
44	101.5	4.3	812	1	IF38_YEAST	P32497 saccharomyc
45	101.5	4.3	2748	1	NIM1_YEAST	Q00402 saccharomyc

RESULT 1
ID YMH1_YEAST STANDARD: PRT: 607 AA.

AC 004632: Q03641:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 69.8 kDa protein in RPL6A-DK1 intergenic region.
GN YMO071C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-113 FROM N.A.
RC STRAIN=5288C / AB972;
RA Bowman S., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RM [2]
RP SEQUENCE OF 88-607 FROM N.A.
RC STRAIN=5288C / AB972;
RA Bowman D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Z38114; CA86249.1; -
DR EMBL: Z46373; CA86507.1; -
DR SGD: S0004536; YMO071C.
KW Hypothetical protein.
SQ SEQUENCE 607 AA; 69828 MW; 57E34E2DE3664B6 CRC64;

Query Match 5.4%; Score 127; DB 1; Length 607;
Best local Similarity 22.1%; Pred. No. 0.28;
Matches 90; Conservative 63; Mismatches 147; Indels 108; Gaps 21;

QY	108	KESDNPANSTKRRFMRGRGNHSYFDLVFNLEKNVTRDADATDI-----EN 158
DB	137	ENDVDGRIMTTSSN---NIAKKKDEDFKALSRLNRSTKEDDKDIDRSITLVLEN 193
QY	159	FAS-----RYLYMATLYYKT-----YTVNDEFGASFNNKLSFT 191
DB	194	IDSITDMLPEPLARTCIPTGHOEAVMLYHTTSLRSFPSTIVDEYCEKVLNLS-T 252
QY	192	TGLFGW-----GIRKALKOIISNPLDIDGTEHSY-----SRLOHTSSYNDY 234
DB	233	TWLSGLVILKLTNVSYSIKITL-OYLSNIPFDKTKNSLSVFLAMKYKRTIDEIASY 311
QY	235	-MDTOIPALPKFAKRFSL--NVVQRLATVAGYVDTPWYKKWMLKLFM--VNRVFITP 289

```

Db      312 PLDVE-----SSNESLLEMMVKRIEVLREHV-----YMSLNVLFKSLPLYDTNDLEIT-   359
Oy      290 KKFENKEIREPSKALKEKYSTDTKDLFEKKIGQVDFPNKEIRDPSSKALKEKVSNDAD   349
Db      360 ---FPEELE--STVIRINOTNEKEELEKE-----KETKREYXQDSVANNED   404
Oy      350 LFEENIGQGVDFPINNEIRDPSSKALIRKXSTGAEDLFEKKIGQVDFPINNEIRDPSSKAL   409
Db      405 VLENK-----SIDVGEV-----QGVVEGSDDAEKTEINEIENYV----NTEEDKAE--   450
Oy      410 IRYVTEADDLFEKKIGQVDFPINKET--RDPSSKALIKVSTEDNL   455
Db      451 --KEKEEEVNTKDNKAKEKEEENINVEVYPEPSPSSIDNKAKKEEKEEI   496

RESULT 2
PRPC_NEIMC
PRPC_NEIMC STANDARD; PRT; 1829 AA.
AC      P55127;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Iron-regulated protein frpc.
GN      FRPC.
OS      Neisseria meningitidis (serogroup C).
OC      Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX      NCBI_TaxID=135720;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=FAM20 / SEROGROUP C.
RX      MEDLINE=94018616; PubMed=8412674;
RA      Thompson S.A., Wang L.L., Sparling P.F.,
RT      Neisseria meningitidis encoding a protein similar to RTX
RT      cytotoxins."
RL      Mol. Microbiol. 9:85-96(1993).
CC      -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
CC      DISEASE.
CC      -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
CC      -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC      CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC      ACTIVITY.
CC      -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L06299; AAA99902.1; .
DR      InterPro; IPR001343; HemIysn_Ca_bind.
DR      Pfam; PF00353; hemIysnCaBind; 9.
DR      PRINTS; PR00313; CABDNNGRPT.
DR      PROSITE; PS00330; HEMOLYSIN_CALCITUM; 17.
KW      Toxin; Calcium; Outer membrane; Repeat.
FT      DOMAIN      879      1702      43 X REPEATS, GLY-RICH.
FT      REPEAT      879      884      1.
FT      REPEAT      888      893      2.
FT      REPEAT      897      902      3.
FT      REPEAT      1016      1021      4.
FT      REPEAT      1025      1030      5.
FT      REPEAT      1034      1039      6.
FT      REPEAT      1043      1048      7.
FT      REPEAT      1052      1057      8.
FT      REPEAT      1061      1066      9.
FT      REPEAT      1070      1075      10.
FT      REPEAT      1079      1084      11.
FT      REPEAT      1088      1093      12.
FT      REPEAT      1097      1102      13.
FT      REPEAT      1216      1221      14.

```

FT	REPEAT	1225	1230	15	
FT	REPEAT	1234	1239	16	
FT	REPEAT	1243	1248	17	
FT	REPEAT	1252	1257	18	
FT	REPEAT	1261	1266	19	
FT	REPEAT	1270	1275	20	
FT	REPEAT	1279	1284	21	
FT	REPEAT	1288	1293	22	
FT	REPEAT	1297	1302	23	
FT	REPEAT	1416	1421	24	
FT	REPEAT	1425	1430	25	
FT	REPEAT	1434	1439	26	
FT	REPEAT	1443	1448	27	
FT	REPEAT	1452	1457	28	
FT	REPEAT	1461	1466	29	
FT	REPEAT	1470	1475	30	
FT	REPEAT	1479	1484	31	
FT	REPEAT	1488	1493	32	
FT	REPEAT	1497	1502	33	
FT	REPEAT	1616	1621	34	
FT	REPEAT	1625	1630	35	
FT	REPEAT	1634	1639	36	
FT	REPEAT	1643	1648	37	
FT	REPEAT	1652	1657	38	
FT	REPEAT	1661	1666	39	
FT	REPEAT	1670	1675	40	
FT	REPEAT	1679	1684	41	
FT	REPEAT	1688	1693	42	
FT	REPEAT	1697	1702	43	
SQ	SEQUENCE	1829 AA;	197622 MW;	5C349020A782DC8 CRC64;	
Query Match 5.2%; Score 122; DB 1; Length 1829;					
Best Local Similarity 20.3%; Pred. No. 2.3; Mismatches 156; Indels 96; Gaps 18;					
Matches 80; Conservative 63; Mismatches 156; Indels 96; Gaps 18;					
QY	94 LPDPOLAEAFILFPESDNPANSTREKPRMRFRGKNHSYFHDLVFNLEKNTVRDA--	151			
DB	35 LQDRADNNAALVIGKDTNLGNL-----LMM--KKG-----VENLMDTVGKKTRL	78			
QY	152 ---DATDIENFASRYLYMATLYKTYTNVDFGASFENKLSFTTGLFGWGIKRALQOIR	208			
DB	79 EKFDVVALQHF--SQYVDLILNKNGHLPNLSIELRSYKAVTY----HGVSSSAALDIVIN	133			
QY	209 SNLP-----LDIGTE---HSVSRLOHTTSYKYDWDIOIPALPK-----FAKRS	250			
DB	134 RSLPDMADGYMALGLGIEAERIHNEQAVNPNPGBSERDNKOLISALDKGFDGSFEKHT	193			
QY	251 LM--VVORLAIYAVAYVDIPWYK--KW-----YMKLNFMVNVFIPTKKFKETREPS	302			
DB	194 FLQSVAMDLTLCGEVETIDGWQKIGGWNGIINDLYKSVKKEWGIPEIYVNNNKKQNE	253			
QY	303 ALKEKVS-----TDTKDLFENKIGQGVDFENKEIRDPSPA	338			
DB	254 AFKNINSILYHMKAAKGEFGDDLNTQNNNLQAAEIIYINDVDTSGIEKGV--AIKE	312			
QY	339 LKEKVSNDAKDLFENKIGQGVDFINNEIRDPSPKALIRKVSATGADLFENKIGQGVDFI	398			
DB	313 LSEKKNASADLADGS-----AEKAKQVDEDLQAQAKEAVEN--AKSTAEKA	357			
QY	399 INNEIRDPSPKAL--IRKVVTEADDLFENKIGQGVDF	431			
DB	358 AQAAKEFFKGLPSFKDLAEKFRDLFPNP--EGWID	390			
RESULT 3					
ID	HEMA_IAUSS	STANDARD;	PRT;	566 AA.	
AC	P03453;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HAZ chain]				

```

Query Match      5.1% Score 121 DB 1 Length 566;
Best Local Similarity 20.8%; Pred. No. 0.63;
Matches 94; Conservative 68; Mismatches 161; Indels 130; Gaps 23.

QY    33 DMSDYLAVS--DNF-----AERICSVPRGSGNSASVAYSRCAKOD-----CLTIQS 80
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    117 ELREQLSVSSEPERETPKER---SWPK-NHVTGGVVASCHCKSSPYRLMLWTEKN 172
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY    81 LKYLEAKYQPLTLPDPIQLAEAFLLFKESDA-----NPANSTSEKKRFMMFFRRGNKS 133
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db 173 GSYF-----NLSSSYNNKKEKELVLMGVHHPSNIEDQKTYR---KENA 214
QY 134 YPHDLVENLLEKNKNTRDADATDIEENFASRYLYMATLYKYKTYNNDEFGASF---FNKLSF 190
Db 215 YVSVSSVNNRRFFPEIERPKRVGOARINYYWTLTIEPGDTLIEFANGNLAPWHAFL 274
QY 191 TTGJFGMGIKRAKLOIISNLPIDIGTEHSVSRLOHTSS--YKQYMDTOIPLPKFKR 248
Db 275 NRG-FGSG-----TITSNMSMDECDTRKCOFPQGAIVSSLPFONIHPTVTIGECPKYVR 326
QY 249 FSLMVY-----ORLLATVAGYVDTPW---YKKMY-----M 275
Db 327 TKLRNVTLGRNIPSIQSGGLFGALNGFIEGGTGMIDGTYGCHHNOEDGSGIADQKSTQ 386
QY 276 KLNEMVNRV-----FIPTRKFPNKIEIPEPSKALKKEKYSTDTKD-----L 315
Db 387 NAINGITRKVSVIEKMKNTQFTAVGKEPNK-LERKREMLNKKVDGDFIDIMTYNAELLVL 445
QY 316 FENKIGQGVDFPFNKETIDPSPKALKEKSNDAKDLFENKIGQGYDF---INNETRDPK 372
Db 446 LEN---ERTLDPEHDSVKNAYEKVKSOLKNNAKE-----IGNGCEPFYHKCNE----- 491
QY 373 ALIRKSTGAEPL---FENKIGQGVTFINNE 401
Db 492 -CMESVKNGTIDYPRYSESKLNREKIDGVKLE 523

RESULT 4
HEMA_I4ZUK
HEMA_I4ZUK STANDARD: PRT: 566 AA.
AC P26141.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HAI chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Swine/Ukkel/1/84).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11517;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Scheil M., Katz J., Kawoka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.";
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INTINATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73775; -; NOT_ANNOTATED_CDS.
DR HSSP: P03437; IRTM.
DR InterPro: IPR001364; Hemaggltn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.

```

```

FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63725 MW; 44661EDB3DB331 CRC64;

Query Match 5.18; Score 120.5; DB 1; Length 566;
Best Local Similarity 19.5%; Pred. No. 0.68;
Matches 107; Conservative 77; Mismatches 186; Indels 179; Gaps 28;

QY 8 GDVYKTLIAESYDSANAMINSDMDIYLSAVS--DNFA-ERICS---QVPRKSGNSCA 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 23 GNSTATYLCIGHAAVPGNTLVKTIYDDQTEVNTATLQVNFSDKICNNPHRILDCANCTL 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 SVSYVMSCAKODCLTLOSLKYPLEAKYQ-----PLTLDPYQVLEAF-----ILFKE 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 83 -IDALL---GPRHCDDEPONEKMDLFVERSKAFNSCYPYDVEDYLSRLASSGLEIFIN 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 SDANPANSTEREFMWRFRGRKNHSYFHDVFNLEK-----NVTADADATDIENFAS 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 EDFMWTGVTQNGSGNACKRGPDSFESRL--NMLYKSGNTYPMLVNT----MPNSDNFK 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 RYLV-----MATLYYKT-----YTNVDEGCA-SFFKSLFTTGLFQWG 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 193 LYINGVHHPTAREQTNLYVQASCKVTVSTRSQOTIIPNIGSRPWAGLSSRIISLY-WT 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 IKR-----ALKQIIRSNPLDIDIGTEHSVRLQHTSS--YK 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 252 IVKQODIIVNSGNLIRPGYFKMHTSRSSIMRSDAPIDICSSSECITPNSIPRDKFQ 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 DYMDQIPLAFKFAKRESLMV-----VORLATVAGYDTPV--YKKWY----- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 312 NVNKITYGACRKYVQNTLKLATGMRNIPKOTRGIFGALINGEWMGVGDFGRH 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 -----MLK-----NEMVNVFPIPTKFEKKEIREPSKAL-----KEKVSFD 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 372 QNSEGTGAADLKSTQAIDQINGKLNVRIEKTNEKFHQIEKESGVGRIQDLEKYED 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 YK-DLFENKIQGVDFENKEIRDPFSKALKEKVSNDADLFENKIQGVDFINNEIRDP 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 432 TKIDLMS-----YNAEL--LVALE-----NQHTIDIDLEEMKML 463
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 SKALIRKVTGAEDLFENKIQGVDFINNEIRDPFSKALIRKVTHEADLFENKIQGV 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 464 FEKTRRQRLRENAEDM-----GNQCF-----KIYHKCDNACTIESIRNGTY 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 DFINKERD 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 503 D--HNEHYD 509
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RX MEDLINE=92317149; PubMed=1618897;
RA Hoyt M.A., He L., Loo K.K., Saunders W.S.;
RT "Two Saccharomyces cerevisiae kinesin-related gene products required
   for mitotic spindle assembly.";
RL J. Cell Biol. 118:109-120(1992).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
   Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
   Mosedale D., Nakahara K., Namath A., Norgren K., Oefner P., Oh C.,
   Petzel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
   Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 1021-1038 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95172238; PubMed=7867803;
RA Rousselet G., Simon M., Ripoch P., Buhler J.M.;
RL FEBS Lett. 359:215-219(1995).
RN [4]
RP CHARACTERIZATION.
RC STRAIN=S288C;
RX MEDLINE=92354062; PubMed=1643659;
RA Saunders W.S., Hoyt M.A.;
RL "Kinesin-related proteins required for structural integrity of the
   mitotic spindle.";
RL Cell 70:451-458(1992).
CC -I- FUNCTION: REQUIRED FOR ASSEMBLY OF THE MITOTIC SPINDLE. INTERACT
   WITH SPINDLE MICROTUBULES TO PRODUCE AN OUTWARDLY DIRECTED
   FORCE ACTING UPON THE POLES. FOLLOWING SPINDLE ASSEMBLY, CIN8 AND
   KIP1 APPARENTLY ACT TO OPPOSE A FORCE THAT DRAMS SEPARATED POLES
   BACK TOGETHER. THIS FORCE SEEMS TO BE MEDIATE BY KAR3.
CC -I- SUBCELLULAR LOCATION: SPINDLE MICROTUBULES THAT LIE BETWEEN THE
   POLES.
CC -I- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BINC
   SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; Z11859; CAAT7885.1; -;
DR EMBL; M90522; AAA34496.1; -;
DR EMBL; U18795; AAB65026.1; -;
DR EMBL; X79105; CAAS5722.1; -;
DR PIR; B42641; B42641.
DR HSSP; P17119; 3KAR.
DR SGD; S0000787; CIN8.
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin.1.
DR PRINTS; PRO0380; KINESINHEAVY.
DR SMART; SMO0129; KISC.1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1.1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2.1.
KW Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
   Mitosis; Cell cycle.
FT DOMAIN 72 553 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 554 657 COILED COIL (POTENTIAL).
FT DOMAIN 904 942 COILED COIL (POTENTIAL).
FT NP_BIND 166 173 ATP (BY SIMILARITY).
FT CONFLICT 254 254 D -> A (IN REF. 1).
FT CONFLICT 831 831 Q -> H (IN REF. 1).
SQ SEQUENCE 1038 AA; 117999 MW; 3A1FD7003EF89FBC CRC64;

Query Match 5.0%; Score 117; DB 1; Length 1038;

```

Best Local Similarity 19.5%; Pred. No. 2.4;
Matches 119; Conservative 97; Mismatches 214; Indels 180; Gaps 28;

```

QY 4 SDSVGVYTLTAAASEVDSANAYMI-----NSDMSDYLSAVSD-----NFA 46
DB 188 SDAAGIIPVLLKLFETLEQNDYVVKVCSFIELYNEELDKLLDSNSGSSNGTGFQGF 247
QY 47 ERICSGVPPGNSASVAVMSRCACQDC-----LTLOS--LKYPLEKKYPTLTPDR 97
DB 248 KLRIRIDSTANTNTNSSSSSSSNSRNSSPRLNDLTPAALLRKLRTKSLPNTIKQ 307
QY 98 YLEAAFIIFKESDANPANSTE-----KRFMRPRRCKNHSYFDLV----- 139
DB 308 YQQQQAIVNSNNSSNSGSGTNNASSNTNNQGRSSMAPNDQNGIYQNLQEPHTNA 367
QY 140 ---FNLEKNVT-RDADADIDENFASR--VLYMATLYKTY-----TN 176
DB 368 MEGLNILQGLKLRHOVAFTKMDPSSRSHTIFITLYKKHODELFRISKANLVLAGSEN 427
QY 177 VDFEG-----ASFENKLSFTTG-----LFGWGKIKRA 202
DB 428 INSGGLNORAKERAGSINOSLLTLGRYINALVDKSGHIPRESKLTRLLODSLGNTKTA 487
QY 203 LKQIIRSNLPIDIGTEHSVSRLOHITSSYKDYMDTQIPALPKFAKPSLM-----V 253
DB 488 L---IATISPAKVTSSETGSTLEY-ASKAKINIK--POLGSFIMDLVKNTMELAKI 541
QY 254 VORLALTAVAG---YVDTPTKKMYMKLKNMVRVPIPTKFFENKIREPSKALKEVST 310
DB 542 KSDLLSTKSEKIGIYMSODHKYKLNLSDESYK-NEV-----OECKREIESITLSKNAL 591
QY 311 DTKDLFENKIGQGVDFENKEIR-----DPSKALKEKYSNDAKDL--FENKIGQGV 359
DB 592 LKVDKLSK---ETIOSQNCQIESLKTTHLPAQLDKOKTELEISDFENKILQKLEVM 648
QY 360 -----VDF-----INNEIRDPKALIRKYSTGAE--LEENKIGQGVDFIN 399
DB 649 QMALHDKRKREDLNQKFEHITKEIKLKSTLFLQNTMQQESILQETNI-QPNIDMIK 707
QY 400 NEIRDPKAL-----IRKYTEADDLFENKIGQGVDFINKEIRDPKAL---I 445
DB 708 NEVLTLRMTQEKALMYKDCVKKILNESKPFNVYIEK--IDIRVDFQKFKNAIENL 765
QY 446 RKVSTEADNL 455
DB 766 SDISENNNA 775

```

RESULT 6
FRPC_NEIMB
ID FRPC_NEIMB STANDARD; PRT; 1829 AA.
AC 09JYV5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpc.
GN FRPC OR NM1415.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RA MEDLINE=20173753; PubMed=10710307;
RA Tetteil H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Ulteback T.R., Khouri H., Qin H., Vamathavan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT *Complete genome sequence of Neisseria meningitidis serogroup B strain

```

RT MC58.":  
RL Science 287:1809-1815(2000).  
CC -I- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL  
CC DISEASE.  
CC -I- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY  
CC SIMILARITY).  
CC -I- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING  
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
CC ACTIVITY.  
CC -I- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: AE002490; AAF41776.1; -.  
DR TIGR: NM1415; -.  
DR InterPro: IPR001343; HemIysn_Ca_bind.  
DR Pfam: PF00353; hemolysinCbind; 9.  
DR PRINTS: PR00313; CABNDNGRPT.  
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 17.  
KW Toxin; Calcium; Outer membrane; Repeat; Complete proteome.  
FT DOMAIN 879 1702  
FT REPEAT 879 884  
FT REPEAT 884 893  
FT REPEAT 893 902  
FT REPEAT 902 1021  
FT REPEAT 1016 1021  
FT REPEAT 1025 1030  
FT REPEAT 1034 1039  
FT REPEAT 1043 1048  
FT REPEAT 1052 1057  
FT REPEAT 1061 1066  
FT REPEAT 1070 1075  
FT REPEAT 1079 1084  
FT REPEAT 1088 1093  
FT REPEAT 1097 1102  
FT REPEAT 1102 1107  
FT REPEAT 1116 1121  
FT REPEAT 1125 1130  
FT REPEAT 1134 1139  
FT REPEAT 1143 1148  
FT REPEAT 1152 1157  
FT REPEAT 1161 1166  
FT REPEAT 1170 1175  
FT REPEAT 1184 1189  
FT REPEAT 1193 1198  
FT REPEAT 1202 1207  
FT REPEAT 1211 1216  
FT REPEAT 1220 1225  
FT REPEAT 1229 1234  
FT REPEAT 1238 1243  
FT REPEAT 1247 1252  
FT REPEAT 1257 1262  
FT REPEAT 1270 1275  
FT REPEAT 1284 1289  
FT REPEAT 1293 1298  
FT REPEAT 1302 1307  
FT REPEAT 1311 1316  
FT REPEAT 1320 1325  
FT REPEAT 1329 1334  
FT REPEAT 1338 1343  
FT REPEAT 1347 1352  
FT REPEAT 1356 1361  
FT REPEAT 1365 1370  
FT REPEAT 1374 1379  
FT REPEAT 1383 1388  
FT REPEAT 1392 1397  
FT REPEAT 1401 1406  
FT REPEAT 1410 1415  
FT REPEAT 1419 1424  
FT REPEAT 1428 1433  
FT REPEAT 1437 1442  
FT REPEAT 1446 1451  
FT REPEAT 1455 1460  
FT REPEAT 1464 1469  
FT REPEAT 1473 1478  
FT REPEAT 1482 1487  
FT REPEAT 1491 1496  
FT REPEAT 1500 1505  
FT REPEAT 1509 1514  
FT REPEAT 1518 1523  
FT REPEAT 1527 1532  
FT REPEAT 1536 1541  
FT REPEAT 1545 1550  
FT REPEAT 1554 1559  
FT REPEAT 1563 1568  
FT REPEAT 1572 1577  
FT REPEAT 1581 1586  
FT REPEAT 1590 1595  
FT REPEAT 1600 1605  
FT REPEAT 1609 1614  
FT REPEAT 1618 1623  
FT REPEAT 1627 1632  
FT REPEAT 1636 1641  
FT REPEAT 1645 1650  
FT REPEAT 1654 1659  
FT REPEAT 1663 1668  
FT REPEAT 1672 1677  
FT REPEAT 1681 1686  
FT REPEAT 1690 1695  
FT REPEAT 1700 1705  
FT REPEAT 1709 1714  
FT REPEAT 1718 1723  
FT REPEAT 1727 1732  
FT REPEAT 1736 1741  
FT REPEAT 1745 1750  
FT REPEAT 1754 1759  
FT REPEAT 1763 1768  
FT REPEAT 1772 1777  
FT REPEAT 1781 1786  
FT REPEAT 1790 1795  
FT REPEAT 1800 1805  
FT REPEAT 1809 1814  
FT REPEAT 1818 1823  
FT REPEAT 1827 1832  
FT REPEAT 1836 1841  
FT REPEAT 1845 1850  
FT REPEAT 1854 1859  
FT REPEAT 1863 1868  
FT REPEAT 1872 1877  
FT REPEAT 1881 1886  
FT REPEAT 1890 1895  
FT REPEAT 1900 1905  
FT REPEAT 1909 1914  
FT REPEAT 1918 1923  
FT REPEAT 1927 1932  
FT REPEAT 1936 1941  
FT REPEAT 1945 1950  
FT REPEAT 1954 1959  
FT REPEAT 1963 1968  
FT REPEAT 1972 1977  
FT REPEAT 1981 1986  
FT REPEAT 1990 1995  
FT REPEAT 2000 2005  
FT REPEAT 2009 2014  
FT REPEAT 2018 2023  
FT REPEAT 2027 2032  
FT REPEAT 2036 2041  
FT REPEAT 2045 2050  
FT REPEAT 2054 2059  
FT REPEAT 2063 2068  
FT REPEAT 2072 2077  
FT REPEAT 2081 2086  
FT REPEAT 2090 2095  
FT REPEAT 2100 2105  
FT REPEAT 2109 2114  
FT REPEAT 2118 2123  
FT REPEAT 2127 2132  
FT REPEAT 2136 2141  
FT REPEAT 2145 2150  
FT REPEAT 2154 2159  
FT REPEAT 2163 2168  
FT REPEAT 2172 2177  
FT REPEAT 2181 2186  
FT REPEAT 2190 2195  
FT REPEAT 2200 2205  
FT REPEAT 2209 2214  
FT REPEAT 2218 2223  
FT REPEAT 2227 2232  
FT REPEAT 2236 2241  
FT REPEAT 2245 2250  
FT REPEAT 2254 2259  
FT REPEAT 2263 2268  
FT REPEAT 2272 2277  
FT REPEAT 2281 2286  
FT REPEAT 2290 2295  
FT REPEAT 2300 2305  
FT REPEAT 2309 2314  
FT REPEAT 2318 2323  
FT REPEAT 2327 2332  
FT REPEAT 2336 2341  
FT REPEAT 2345 2350  
FT REPEAT 2354 2359  
FT REPEAT 2363 2368  
FT REPEAT 2372 2377  
FT REPEAT 2381 2386  
FT REPEAT 2390 2395  
FT REPEAT 2400 2405  
FT REPEAT 2409 2414  
FT REPEAT 2418 2423  
FT REPEAT 2427 2432  
FT REPEAT 2436 2441  
FT REPEAT 2445 2450  
FT REPEAT 2454 2459  
FT REPEAT 2463 2468  
FT REPEAT 2472 2477  
FT REPEAT 2481 2486  
FT REPEAT 2490 2495  
FT REPEAT 2500 2505  
FT REPEAT 2509 2514  
FT REPEAT 2518 2523  
FT REPEAT 2527 2532  
FT REPEAT 2536 2541  
FT REPEAT 2545 2550  
FT REPEAT 2554 2559  
FT REPEAT 2563 2568  
FT REPEAT 2572 2577  
FT REPEAT 2581 2586  
FT REPEAT 2590 2595  
FT REPEAT 2600 2605  
FT REPEAT 2609 2614  
FT REPEAT 2618 2623  
FT REPEAT 2627 2632  
FT REPEAT 2636 2641  
FT REPEAT 2645 2650  
FT REPEAT 2654 2659  
FT REPEAT 2663 2668  
FT REPEAT 2672 2677  
FT REPEAT 2681 2686  
FT REPEAT 2690 2695  
FT REPEAT 2700 2705  
FT REPEAT 2709 2714  
FT REPEAT 2718 2723  
FT REPEAT 2727 2732  
FT REPEAT 2736 2741  
FT REPEAT 2745 2750  
FT REPEAT 2754 2759  
FT REPEAT 2763 2768  
FT REPEAT 2772 2777  
FT REPEAT 2781 2786  
FT REPEAT 2790 2795  
FT REPEAT 2800 2805  
FT REPEAT 2809 2814  
FT REPEAT 2818 2823  
FT REPEAT 2827 2832  
FT REPEAT 2836 2841  
FT REPEAT 2845 2850  
FT REPEAT 2854 2859  
FT REPEAT 2863 2868  
FT REPEAT 2872 2877  
FT REPEAT 2881 2886  
FT REPEAT 2890 2895  
FT REPEAT 2900 2905  
FT REPEAT 2909 2914  
FT REPEAT 2918 2923  
FT REPEAT 2927 2932  
FT REPEAT 2936 2941  
FT REPEAT 2945 2950  
FT REPEAT 2954 2959  
FT REPEAT 2963 2968  
FT REPEAT 2972 2977  
FT REPEAT 2981 2986  
FT REPEAT 2990 2995  
FT REPEAT 3000 3005  
FT REPEAT 3009 3014  
FT REPEAT 3018 3023  
FT REPEAT 3027 3032  
FT REPEAT 3036 3041  
FT REPEAT 3045 3050  
FT REPEAT 3054 3059  
FT REPEAT 3063 3068  
FT REPEAT 3072 3077  
FT REPEAT 3081 3086  
FT REPEAT 3090 3095  
FT REPEAT 3100 3105  
FT REPEAT 3109 3114  
FT REPEAT 3118 3123  
FT REPEAT 3127 3132  
FT REPEAT 3136 3141  
FT REPEAT 3145 3150  
FT REPEAT 3154 3159  
FT REPEAT 3163 3168  
FT REPEAT 3172 3177  
FT REPEAT 3181 3186  
FT REPEAT 3190 3195  
FT REPEAT 3200 3205  
FT REPEAT 3209 3214  
FT REPEAT 3218 3223  
FT REPEAT 3227 3232  
FT REPEAT 3236 3241  
FT REPEAT 3245 3250  
FT REPEAT 3254 3259  
FT REPEAT 3263 3268  
FT REPEAT 3272 3277  
FT REPEAT 3281 3286  
FT REPEAT 3290 3295  
FT REPEAT 3300 3305  
FT REPEAT 3309 3314  
FT REPEAT 3318 3323  
FT REPEAT 3327 3332  
FT REPEAT 3336 3341  
FT REPEAT 3345 3350  
FT REPEAT 3354 3359  
FT REPEAT 3363 3368  
FT REPEAT 3372 3377  
FT REPEAT 3381 3386  
FT REPEAT 3390 3395  
FT REPEAT 3400 3405  
FT REPEAT 3409 3414  
FT REPEAT 3418 3423  
FT REPEAT 3427 3432  
FT REPEAT 3436 3441  
FT REPEAT 3445 3450  
FT REPEAT 3454 3459  
FT REPEAT 3463 3468  
FT REPEAT 3472 3477  
FT REPEAT 3481 3486  
FT REPEAT 3490 3495  
FT REPEAT 3500 3505  
FT REPEAT 3509 3514  
FT REPEAT 3518 3523  
FT REPEAT 3527 3532  
FT REPEAT 3536 3541  
FT REPEAT 3545 3550  
FT REPEAT 3554 3559  
FT REPEAT 3563 3568  
FT REPEAT 3572 3577  
FT REPEAT 3581 3586  
FT REPEAT 3590 3595  
FT REPEAT 3600 3605  
FT REPEAT 3609 3614  
FT REPEAT 3618 3623  
FT REPEAT 3627 3632  
FT REPEAT 3636 3641  
FT REPEAT 3645 3650  
FT REPEAT 3654 3659  
FT REPEAT 3663 3668  
FT REPEAT 3672 3677  
FT REPEAT 3681 3686  
FT REPEAT 3690 3695  
FT REPEAT 3700 3705  
FT REPEAT 3709 3714  
FT REPEAT 3718 3723  
FT REPEAT 3727 3732  
FT REPEAT 3736 3741  
FT REPEAT 3745 3750  
FT REPEAT 3754 3759  
FT REPEAT 3763 3768  
FT REPEAT 3772 3777  
FT REPEAT 3781 3786  
FT REPEAT 3790 3795  
FT REPEAT 3800 3805  
FT REPEAT 3809 3814  
FT REPEAT 3818 3823  
FT REPEAT 3827 3832  
FT REPEAT 3836 3841  
FT REPEAT 3845 3850  
FT REPEAT 3854 3859  
FT REPEAT 3863 3868  
FT REPEAT 3872 3877  
FT REPEAT 3881 3886  
FT REPEAT 3890 3895  
FT REPEAT 3900 3905  
FT REPEAT 3909 3914  
FT REPEAT 3918 3923  
FT REPEAT 3927 3932  
FT REPEAT 3936 3941  
FT REPEAT 3945 3950  
FT REPEAT 3954 3959  
FT REPEAT 3963 3968  
FT REPEAT 3972 3977  
FT REPEAT 3981 3986  
FT REPEAT 3990 3995  
FT REPEAT 4000 4005  
FT REPEAT 4009 4014  
FT REPEAT 4018 4023  
FT REPEAT 4027 4032  
FT REPEAT 4036 4041  
FT REPEAT 4045 4050  
FT REPEAT 4054 4059  
FT REPEAT 4063 4068  
FT REPEAT 4072 4077  
FT REPEAT 4081 4086  
FT REPEAT 4090 4095  
FT REPEAT 4100 4105  
FT REPEAT 4109 4114  
FT REPEAT 4118 4123  
FT REPEAT 4127 4132  
FT REPEAT 4136 4141  
FT REPEAT 4145 4150  
FT REPEAT 4154 4159  
FT REPEAT 4163 4168  
FT REPEAT 4172 4177  
FT REPEAT 4181 4186  
FT REPEAT 4190 4195  
FT REPEAT 4200 4205  
FT REPEAT 4209 4214  
FT REPEAT 4218 4223  
FT REPEAT 4227 4232  
FT REPEAT 4236 4241  
FT REPEAT 4245 4250  
FT REPEAT 4254 4259  
FT REPEAT 4263 4268  
FT REPEAT 4272 4277  
FT REPEAT 4281 4286  
FT REPEAT 4290 4295  
FT REPEAT 4300 4305  
FT REPEAT 4309 4314  
FT REPEAT 4318 4323  
FT REPEAT 4327 4332  
FT REPEAT 4336 4341  
FT REPEAT 4345 4350  
FT REPEAT 4354 4359  
FT REPEAT 4363 4368  
FT REPEAT 4372 4377  
FT REPEAT 4381 4386  
FT REPEAT 4390 4395  
FT REPEAT 4400 4405  
FT REPEAT 4409 4414  
FT REPEAT 4418 4423  
FT REPEAT 4427 4432  
FT REPEAT 4436 4441  
FT REPEAT 4445 4450  
FT REPEAT 4454 4459  
FT REPEAT 4463 4468  
FT REPEAT 4472 4477  
FT REPEAT 4481 4486  
FT REPEAT 4490 4495  
FT REPEAT 4500 4505  
FT REPEAT 4509 4514  
FT REPEAT 4518 4523  
FT REPEAT 4527 4532  
FT REPEAT 4536 4541  
FT REPEAT 4545 4550  
FT REPEAT 4554 4559  
FT REPEAT 4563 4568  
FT REPEAT 4572 4577  
FT REPEAT 4581 4586  
FT REPEAT 4590 4595  
FT REPEAT 4600 4605  
FT REPEAT 4609 4614  
FT REPEAT 4618 4623  
FT REPEAT 4627 4632  
FT REPEAT 4636 4641  
FT REPEAT 4645 4650  
FT REPEAT 4654 4659  
FT REPEAT 4663 4668  
FT REPEAT 4672 4677  
FT REPEAT 4681 4686  
FT REPEAT 4690 4695  
FT REPEAT 4700 4705  
FT REPEAT 4709 4714  
FT REPEAT 4718 4723  
FT REPEAT 4727 4732  
FT REPEAT 4736 4741  
FT REPEAT 4745 4750  
FT REPEAT 4754 4759  
FT REPEAT 4763 4768  
FT REPEAT 4772 4777  
FT REPEAT 4781 4786  
FT REPEAT 4790 4795  
FT REPEAT 4800 4805  
FT REPEAT 4809 4814  
FT REPEAT 4818 4823  
FT REPEAT 4827 4832  
FT REPEAT 4836 4841  
FT REPEAT 4845 4850  
FT REPEAT 4854 4859  
FT REPEAT 4863 4868  
FT REPEAT 4872 4877  
FT REPEAT 4881 4886  
FT REPEAT 4890 4895  
FT REPEAT 4900 4905  
FT REPEAT 4909 4914  
FT REPEAT 4918 4923  
FT REPEAT 4927 4932  
FT REPEAT 4936 4941  
FT REPEAT 4945 4950  
FT REPEAT 4954 4959  
FT REPEAT 4963 4968  
FT REPEAT 4972 4977  
FT REPEAT 4981 4986  
FT REPEAT 4990 4995  
FT REPEAT 5000 5005  
FT REPEAT 5009 5014  
FT REPEAT 5018 5023  
FT REPEAT 5027 5032  
FT REPEAT 5036 5041  
FT REPEAT 5045 5050  
FT REPEAT 5054 5059  
FT REPEAT 5063 5068  
FT REPEAT 5072 5077  
FT REPEAT 5081 5086  
FT REPEAT 5090 5095  
FT REPEAT 5100 5105  
FT REPEAT 5109 5114  
FT REPEAT 5118 5123  
FT REPEAT 5127 5132  
FT REPEAT 5136 5141  
FT REPEAT 5145 5150  
FT REPEAT 5154 5159  
FT REPEAT 5163 5168  
FT REPEAT 5172 5177  
FT REPEAT 5181 5186  
FT REPEAT 5190 5195  
FT REPEAT 5200 5205  
FT REPEAT 5209 5214  
FT REPEAT 5218 5223  
FT REPEAT 5227 5232  
FT REPEAT 5236 5241  
FT REPEAT 5245 5250  
FT REPEAT 5254 5259  
FT REPEAT 5263 5268  
FT REPEAT 5272 5277  
FT REPEAT 5281 5286  
FT REPEAT 5290 5295  
FT REPEAT 5300 5305  
FT REPEAT 5309 5314  
FT REPEAT 5318 5323  
FT REPEAT 5327 5332  
FT REPEAT 5336 5341  
FT REPEAT 5345 5350  
FT REPEAT 5354 5359  
FT REPEAT 5363 5368  
FT REPEAT 5372 5377  
FT REPEAT 5381 5386  
FT REPEAT 5390 5395  
FT REPEAT 5400 5405  
FT REPEAT 5409 5414  
FT REPEAT 5418 5423  
FT REPEAT 5427 5432  
FT REPEAT 5436 5441  
FT REPEAT 5445 5450  
FT REPEAT 5454 5459  
FT REPEAT 5463 5468  
FT REPEAT 5472 5477  
FT REPEAT 5481 5486  
FT REPEAT 5490 5495  
FT REPEAT 5500 5505  
FT REPEAT 5509 5514  
FT REPEAT 5518 5523  
FT REPEAT 5527 5532  
FT REPEAT 5536 5541  
FT REPEAT 5545 5550  
FT REPEAT 5554 5559  
FT REPEAT 5563 5568  
FT REPEAT 5572 5577  
FT REPEAT 5581 5586  
FT REPEAT 5590 5595  
FT REPEAT 5600 5605  
FT REPEAT 5609 5614  
FT REPEAT 5618 5623  
FT REPEAT 5627 5632  
FT REPEAT 5636 5641  
FT REPEAT 5645 5650  
FT REPEAT 5654 5659  
FT REPEAT 5663 5668  
FT REPEAT 5672 5677  
FT REPEAT 5681 5686  
FT REPEAT 5690 5695  
FT REPEAT 5700 5705  
FT REPEAT 5709 5714  
FT REPEAT 5718 5723  
FT REPEAT 5727 5732  
FT REPEAT 5736 5741  
FT REPEAT 5745 5750  
FT REPEAT 5754 5759  
FT REPEAT 5763 5768  
FT REPEAT 5772 5777  
FT REPEAT 5781 5786  
FT REPEAT 5790 5795  
FT REPEAT 5800 5805  
FT REPEAT 5809 5814  
FT REPEAT 5818 5823  
FT REPEAT 5827 5832  
FT REPEAT 5836 5841  
FT REPEAT 5845 5850  
FT REPEAT 5854 5859  
FT REPEAT 5863 5868  
FT REPEAT 5872 5877  
FT REPEAT 5881 5886  
FT REPEAT 5890 5895  
FT REPEAT 5900 5905  
FT REPEAT 5909 5914  
FT REPEAT 5918 5923  
FT REPEAT 5927 5932  
FT REPEAT 5936 5941  
FT REPEAT 5945 5950  
FT REPEAT 5954 5959  
FT REPEAT 5963 5968  
FT REPEAT 5972 5977  
FT REPEAT 5981 5986  
FT REPEAT 5990 5995  
FT REPEAT 6000 6005  
FT REPEAT 6009 6014  
FT REPEAT 6018 6023  
FT REPEAT 6027 6032  
FT REPEAT 6036 6041  
FT REPEAT 6045 6050  
FT REPEAT 6054 6059  
FT REPEAT 6063 6068  
FT REPEAT 6072 6077  
FT REPEAT 6081 6086  
FT REPEAT 6090 6095  
FT REPEAT 6100 6105  
FT REPEAT 6109 6114  
FT REPEAT 6118 6123  
FT REPEAT 6127 6132  
FT REPEAT 6136 6141  
FT REPEAT 6145 6150  
FT REPEAT 6154 6159  
FT REPEAT 6163 6168  
FT REPEAT 6172 6177  
FT REPEAT 6181 6186  
FT REPEAT 6190 6195  
FT REPEAT 6200 6205  
FT REPEAT 6209 6214  
FT REPEAT 6218 6223  
FT REPEAT 6227 6232  
FT REPEAT 6236 6241  
FT REPEAT 6245 6250  
FT REPEAT 6254 6259  
FT REPEAT 6263 6268  
FT REPEAT 6272 6277  
FT REPEAT 6281 6286  
FT REPEAT 6290 6295  
FT REPEAT 6300 6305  
FT REPEAT 6309 6314  
FT REPEAT 6318 6323  
FT REPEAT 6327 6332  
FT REPEAT 6336 6341  
FT REPEAT 6345 6350  
FT REPEAT 6354 6359  
FT REPEAT 6363 6368  
FT REPEAT 6372 6377  
FT REPEAT 6381 6386  
FT REPEAT 6390 6395  
FT REPEAT 6400 6405  
FT REPEAT 6409 6414  
FT REPEAT 6418 6423  
FT REPEAT 6427 6432  
FT REPEAT 6436 6441  
FT REPEAT 6445 6450  
FT REPEAT 6454 6459  
FT REPEAT 6463 6468  
FT REPEAT 6472 6477  
FT REPEAT 6481 6486  
FT REPEAT 6490 6495  
FT REPEAT 6500 6505  
FT REPEAT 6509 6514  
FT REPEAT 6518 6523  
FT REPEAT 6527 6532  
FT REPEAT 6536 6541  
FT REPEAT 6545 6550  
FT REPEAT 6554 6559  
FT REPEAT 6563 6568  
FT REPEAT 6572 6577  
FT REPEAT 6581 6586  
FT REPEAT 6590 6595  
FT REPEAT 6600 6605  
FT REPEAT 6609 6614  
FT REPEAT 6618 6623  
FT REPEAT 6627 6632  
FT REPEAT 6636 6641  
FT REPEAT 6645 6650  
FT REPEAT 6654 6659  
FT REPEAT 6663 6668  
FT REPEAT 6672 6677  
FT REPEAT 6681 6686  
FT REPEAT 6690 6695  
FT REPEAT 6700 6705  
FT REPEAT 6709 6714  
FT REPEAT 6718 6723  
FT REPEAT 6727 6732  
FT REPEAT 6736 6741  
FT REPEAT 6745 6750  
FT REPEAT 6754 6759  
FT REPEAT 6763 6768  
FT REPEAT 6772 6777  
FT REPEAT 6781 6786  
FT REPEAT 6790 6795  
FT REPEAT 6800 6805  
FT REPEAT 6809 6814  
FT REPEAT 6818 6823  
FT REPEAT 6827 6832  
FT REPEAT 6836 6841  
FT REPEAT 6845 6850  
FT REPEAT 6854 6859  
FT REPEAT 6863 6868  
FT REPEAT 6872 6877  
FT REPEAT 6881 6886  
FT REPEAT 6890 6895  
FT REPEAT 6900 6905  
FT REPEAT 6909 6914  
FT REPEAT 6918 6923  
FT REPEAT 6927 6932  
FT REPEAT 6936 6941  
FT REPEAT 6945 6950  
FT REPEAT 6954 6959  
FT REPEAT 6963 6968  
FT REPEAT 6972 6977  
FT REPEAT 6981 6986  
FT REPEAT 6990 6995  
FT REPEAT 7000 7005  
FT REPEAT 7009 7014  
FT REPEAT 7018 7023  
FT REPEAT 7027 7032  
FT REPEAT 7036 7041  
FT REPEAT 7045 7050  
FT REPEAT 7054 7059  
FT REPEAT 7063 7068  
FT REPEAT 7072 7077  
FT REPEAT 7081 7086  
FT REPEAT 7090 7095  
FT REPEAT 7100 7105  
FT REPEAT 7109 7114  
FT REPEAT 7118 7123  
FT REPEAT 7127 7132  
FT REPEAT 7136 7141  
FT REPEAT 7145 7150  
FT REPEAT 7154 7159  
FT REPEAT 7163 7168  
FT REPEAT 7172 7177  
FT REPEAT 7181 7186  
FT REPEAT 7190 7195  
FT REPEAT 7200 7205  
FT REPEAT 7209 7214  
FT REPEAT 7218 7223  
FT REPEAT 7227 7232  
FT REPEAT 7236 72
```

```

Best Local Similarity 20.2%, Pred. No. 4.9:
Matches 81: Conservative 62: Mismatches 150: Indels 108: Gaps 19:

QY      94 LPPYQLEAAFLLEKESDANPANSTEKRFMRFRGKHNHYPHDLVFNLLEKNVTRDADA 153
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      35 LGDRVADDDAAALVGKDANLNGLN-----LMM--KKG-----VENLMDTVK--K 75

QY      154 TDIEAFASRYLMAFLTYKTYT-----NDFEGASFENKLSFTTGLFGMGIKRA 202
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      76 TRLEKFDR---VALQHFROYARLINONNGRLPNTSEIERSYKAVTDN---GVSSSA 127

QY      203 LKQIIRSNLP-----LDIGTE-----HSVSRGLHITSYVDWMDTQIPALPK----- 244
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      128 IDLVNRSRPLMDADGYMALGLGIERHINEQAVNPNPGESEDNKRKQLTSALDKFDSGF 187

QY      245 FAKRFSLMVYQRLATVAG--YVDTPMYK--KW-----YMKLKNFVNRVFIPTKKFPNKE 296
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      168 KEKHTEFLQSVAMDVTKLGVEYITDGWOKIGWGNGIINDLYKSSVYKKRWGTFIFELVYNN 247

QY      297 IREPSALKLEKYS-----TDTKDLFENKIGOGTVDFENKEI 332
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      248 IKQGEARKNNEINSLVHDMKAAGREFGDDLNTQNMNNLQQAELIINDYDNTSGQIEKV 307

QY      333 RDPSPALKLEKYSNDAKDLFENKIGOGTVDFINNEIRDSPALKIRKYSTGAEDLFENKIGQ 392
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      308 K-AIKELSEKMNNAASDLADGS-----AEKAKYVEDDLQAQAKPAYEN--AK 351

QY      393 GTVDFINNEIRDSPAL--IRKYVTEADDLFENKIGOGTVD 431
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      352 STAEKAAQAAREFFGLPSFKDLAEKFDRLFPNP--EGWID 390

```

```

RESULT 7
Y162_UREPA          STANDARD:      PRT:      457 AA.
ID   Y162_UREPA
AC   Q9POY2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein u0162.
GN   u0162.
OS   Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Ureaplasma.
OX   NCBI_TaxID=134821;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEKOVAR 3;
RX   MEDLINE=20500219; PubMed=11048724;
RA   Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA   Cassell G.H.;
RT   "The complete sequence of the mucosal pathogen Ureaplasma
RL   Nature 407:757-762(2000).
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb.ch/announce/isb.ch).
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; AE002116; AAF30568.1; -
DR   InterPro; IPR000087; Collagen.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 457 AA; 51904 MW; 57D0EFC8D0A4D637 CRC64;
Query Match          4.9%; Score 116; DB 1; Length 457;
Best Local Similarity 19.4%; Pred. No. 1;
Matches      83; Conservative      61; Mismatches 133; Indels 150; Gaps 19;

```

```

Db 25 LEDNOLQLEKRRQAL-----TFLRAQSMIDCVASNG-----GAGFIYSW 66
QY 135 FHDVFNLEKKNVTRDADAT-----154
67 FDDLTITDYQKTRFEALATLTNTLMLQGHFIDKQYTSALIGDGLSLSPKSENADEPKRL 126
QY 155 DIENFASRLYMATLYYKT-----YTVNDEFGASFFFKLSFTTGLGSGWGIKRALK 204
127 DIIN-----KLRFQGLYKSKSFANENDDGIYUVDELKINAIN-----NVHEELR 171
QY 205 Q-----IIRSNPLDIGFHSYRLOHTSS-----YKQYM--DPOIP--ALPKFA 246
172 DHIYKKNPVPVANNNDLN---HHNNINNVSSINASGNLATLNGFILTNDCTLPPNDLTTDDSN 228
QY 247 KRESLWVVQRLLATYAGYVDFPWWK--WYMKLKNPVNRYVTPKTFEKNFIEIREPSAL 304
229 KYVD-----NAIANTYVDIHQVKNLAWYGNNNNAITR-----OELYEYCKI 270
QY 305 KEKVSTDRKDFENKIGOGTVDFENK-----EIRDSKALKEKVS- 345
271 NSKMWDA--IYDYAGELVYDLDEKTYIKYKALVNVNGVSTHNVDPSPKINENWEL 328
QY 346 -DAKDLFENKIGOGTVDFINNEIRDSKALRKVSTGAEDLEFNKIGOGTVDFINNEIRD 404
Db 349 INIQIDLNKIKEGVQDAVDNKNVENVKRAIDTKENTFTKNL-ENKVNBEAKQAVLSSYED 387
QY 405 PSKALIR 411
Db 388 KASGVFK 394

```

```

CC RESULT 8
CC MLPL_YEAST
CC ID MLPL_YEAST STANDARD: PRT: 1875 AA.
CC AC Q02455:
CC DT 01-OCT-1993 (Rel. 27, Created)
CC DT 01-JUN-1994 (Rel. 29, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Myosin-like protein MLPL1.
CC GN MLPL OR YKR055W OR YKR415.
CC OS Saccharomyces cerevisiae (Baker's yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CC OX NCBI_TaxID=4932;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=S288C;
CC RX MEDLINE=93247549; PubMed=8483450;
CC RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
CC RT "A new yeast gene with a myosin-like heptad repeat structure.";
CC RL Mol. Gen. Genet. 237:359-369(1993).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=94205265; PubMed=8154186;
CC RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
CC RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
CC RT "The complete sequence of a 15,820 bp segment of Saccharomyces
CC RT cerevisiae chromosome XI contains the UBI2 and MLPL genes and three
CC RT new open reading frames.";
CC RL Yeast 9:1349-1354(1993).
CC CC
CC -I- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC
CC -I- SIMILARITY: SOME. TO THE TPR ONCOGENE.
CC -I- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC CC or send an email to license@isb-sib.ch).

```


CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDPA ATPASE
CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CIS AT THE SH-1
CC POSITION (688).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M14628: AAA3227.1: -
DR PIR: A26655: A26655.
DR PIR: S00250: S00250.
DR PDB: 1MMA: 03-DEC-97.
DR PDB: 1MMD: 17-AUG-96.
DR PDB: 1MMG: 03-DEC-97.
DR PDB: 1MMN: 03-DEC-97.
DR PDB: 1MND: 17-AUG-96.
DR PDB: 1MNE: 17-AUG-96.
DR PDB: 1VOM: 23-DEC-96.
DR PDB: 1LVK: 28-JAN-98.
DR DICTYD: DD01008: mbca.
DR InterPro: IPR000048: IQ.
DR InterPro: IPR004009: Myosin_N.
DR InterPro: IPR001609: myosin_head.
DR Pfam: PF00612: IQ: 2.
DR Pfam: PF00063: myosin_head: 1.
DR Pfam: PF02736: Myosin_N: 1.
DR PRINTS: PR00193: MYOSINHEAVY.
DR PRODOM: PD000355: myosin_head: 1.
DR SMART: SM00015: IQ: 1.
DR SMART: SM00242: MYSC: 1.
DR Prosite: PS00096: IQ: 1.
DR Myosin: Colled coil: Actin-binding: ATP-binding: 3D-structure:
KW Calmodulin-binding: Methylation: Alkylation: Phosphorylation.
FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
FT DOMAIN 762 791 IQ.
FT DOMAIN 817 2116 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP.
FT DOMAIN 638 660 ACTIN-BINDING.
FT DOMAIN 738 752 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
FT MOD_RES 678 678 ALKYLATION (SH-1).
FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
SQ SEQUENCE 2116 AA; 243871 MW; 2FC370BBI5E56A1 CRC64;
Query Match 4.8%; Score 113.5; DB 1; Length 2116;
Best Local Similarity 20.0%; Pred. No. 10;
Matches 91; Conservative 77; Mismatches 166; Indels 121; Gaps 23;
OY 42 SNFMR-----ICQVPGKSCASVASAYKRCACODCLTLOSILKYLEAKYOPTLP 95
DB 604 SONVVTKLENDNPDIASRAKKANFTVAQY-----KEQLASLMTLETNTNHPFRCTIP 658
OY 96 D-----PYOLEAFILFKESDANPANSTERKFRFRKGNHSYFDLV--FNILEKNVTR 149
DB 659 NKKQIPAKIEDVVL-----DQLRCNGVLEGITRKGFNRKRIITYADFKRYILLAPNVR 714
OY 150 DAD-----ATDIENFASRYLYMATLYKYTYTND-----EFGASFNKLSTFTGIFGCIKR 201
DB 715 DAEDSOKATD-----AVL-----KHLNIDPEQYRG--ITRKIFRAGQLA-RIEE 756
OY 202 ALKOIRISLU-PLDIGTEHSVRLQHTITSSYKYDYMDTQIPALPKRAKRSLSNVQRLAT 260
DB 757 AREORISEIQAIAQATRWIAR-----KVYQAREHTVAA-----RIIDONLRA 801

OY 261 VAGYDTPYKMKWYKLNKFNVRVETPKFENKEIREPSKAL-----REK 307
DB 802 YIDFSWPMW-KLFSKARLL-----KRNFETIKKEEILELKNLSDSTQOK 853
OY 308 VSTDTRKDFENKIGOGTVDFPNKEIRPSKALKE--KVSNDARDFENKIGOGTVDFINN 365
DB 854 LEKSLKDTRESNVL-----DLQRLKAEKETLKAAYDSKDALE----- 890
OY 366 EIRDPKALIRKVSIGAEEDLFENKIGOGTVDFINNIRDPKALIRKYTEADDFENKI 425
DB 891 -----AKRELEIRVEQM-BSEIDCKLALEN--LQOKRSVEKVRDLEELQEEOK 940
OY 426 GOGTVDFINKEIRDPKALIRKVSIEDADNL--LEK 458
DB 941 LNTLEKLKKYEELEEMKRVNDGSDITSRLK 975
RESULT 10
MAPB_RAT STANDARD: PRT; 2459 AA.
AC P15205: 062958: Q9ER21: Q9QW92:
DT 01-APR-1990 (rel. 14, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
RA MEDLINE=96257242; PubMed=8666295;
RX Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain; and Gliat tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Felck P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rlenitz A., Greeningloh G., Hermans-Borgmeyer I., Kirsch J.,
Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -1- FUNCTION: The function of brain MAPs is essentially unknown.
CC phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.

FT CHAIN 346 566 HEMAGGLUTININ HAZ CHAIN.
 FT CAROHD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 566 AA: 63514 MW: AEC60738C765F637 CRC64;

Query Match 4.7%: Score 112; DB 1; Length 566;
 Best Local Similarity 20.5%; Pred. No. 2.5;
 Matches 93; Conservative 68; Mismatches 162; Indels 130; Gaps 23;

QY 33 DMSDYLAVS--DNF---AERICSOVPGKSNCSASVAYMSRCADOD-----CITLOS 80
 DB 117 ELREOLSVSVSEFPEIFPKER---SWPK-HNYTRGVTASCSSHKGSFYNILMLTEKN 172
 QY 81 LKYPLEAKYQPLTLPDPYQLEAFLFKESDA-----NPANSTPKRRMRRGRKNHS 133
 DB 173 GSP-----NLKSYVNNKEKEVLVMGHHPSNIEDQKTYR---KENA 214
 QY 134 YPHDLVNLLEKNVTRDADATDIENFASRYLYMATLYKTYTNVDEFGASF--FNKLSF 190
 DB 215 YSVSVSNVNRFTPEIARPKVYGAGRIINYWTLLERGDITIFANONLILAPMAFL 274
 QY 191 TGLGLGWMGKRLKQIIRNLPDIDGTEHSVSLQHTSS--YKDYMDIQIPALPFAKR 248
 DB 275 NRG-EGSG-----ITISNASMDECDTCKQTPOGAINSLSFONHPYITGECPRYVS 326
 QY 249 FSLMVV-----ORLLATVAGYVDTPW---YKKVY-----M 275
 DB 327 TILRVATGLRNIPSTOSRGLFGAIGFICGWTGMIDGVIYNHNOEGSSVAAADOKSQ 386
 QY 276 KUKNFMVNV-----FIPYKKFKNKEIREPSKALKKYSTDTKD-----L 315
 DB 387 NAINGITKNVNSVIEKMTQFTAVGKEFD-ELEKRMENLNKKVDOGFLDIWYNAELVL 445
 QY 316 FENKIGOGVDFPNKEIRPSPKALKKESVNDADLFEKNIKIGQVDF--INNEIDPSK 372
 DB 446 LEN---ERLIDPHDSNVKMLYEKVSQLEKNNAKE---IGNGCEFFYHKCNNE----- 491
 QY 373 ALIRKYSTGAEDL---FENKIGOGVDFINNE 401
 DB 492 -CMESYKNGTYDYPKYSKSLNREKIDGKLE 523

RESULT 12
 YD86_SCHPO STANDARD: PRT: 1957 AA.
 AC 010411:
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE Hypothetical 222.8 kDa protein Clp3.06C in chromosome I.
 GN SPCLF3.06C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: 270690; CAA94624.1; .
 KW Hypothetical protein.
 SQ SEQUENCE 1957 AA: 222785 MW: 3F480CA06171D9DA CRC64;

Query Match 4.7%: Score 112; DB 1; Length 1957;
 Best Local Similarity 19.2%; Pred. No. 11;
 Matches 107; Conservative 94; Mismatches 204; Indels 152; Gaps 27;

QY 9 DWTKTLAASEVDSAAANY-MINSDMSDYLAVSNDFA-----BRICSOVPGKSN 58
 DB 205 DLSROLTYTERKDKKEXYERKEDVSSIKSLAEOGNSNLSRGQRL-EKLLVSN 263
 QY 59 CSASVAYMSRCADQDLTLQ-----SLKYPLEAKYQPLTLPDPYQLE 101
 DB 264 KTVSTLRQTENSLRACCKTLQEKLEKCAINEDSKLLEELKHNV-ANYSDAIVHKDLIE 322
 QY 102 AAFLEKESDANPANTERKFRMRFRRC-----NHSYFDLVNMLE 144
 DB 323 DISTRISEPD---NLKSRDITLSTINEKLEKLRNTIGSLKDSRTSNQLEEMVELKE 378
 QY 145 KNYTRDADATDIENFASRYLYMATLYKTYTNVDEFGASFNKLSPFTGLFGWGIKRLK 204
 DB 379 SNRTINSQLTDAESKLSF---EQEKSLKGSIDEVQ---NLSKSKDM---VKQVSS 427
 QY 205 QI--IRSNLPDIDGTEHSVSLQHTSSYKDYMDIQIPALPFAK--FSLVAYORLLAT 260
 DB 428 QLEEARSSL-----AHAAGKLAELINSE-RDFQNKIKIKFEIETDRLACLMSNELKE 480
 QY 261 VAGYVDPYKWKYMKLKNPVMNRVFIPTKKFKNKEIREPSKALKKYSTDTK--OLEFN 318
 DB 481 KSLALD-----KKDOELNLN---REQIEQKVSSESTQSSLSQSLQDLITNEKKKHEVYS 532
 QY 319 KI-----GOGVDFENKE-----IRPSPALKKESVND 345
 DB 533 QINELKGLQTELSNEHLSQSLTSLAEKPAVAATNNELSSKNSLQTLCAVFOKRLAK 592
 QY 346 DAKDLFEN-----KIGOGVDFINN-----EIRPSPALKKYSTGAEDLFE 387
 DB 593 SYWQLEKNQNFSSLDTSFKKLINESHOELNNHQTITKOLKPTSSKLOLOLERAN--FE 650
 QY 388 NIGOGVDFINNEIR-----DPSKALIRKYVTEADDLFENKIGOGVDFINKEIRDP 440
 DB 651 QK--ESTLSDENNDLTKLLKLEESNKSILIK-QEDVDSLERN-----TQTLKEDLRS 701
 QY 441 SKALIRKYSTEADNMLE 457
 DB 702 EEARL-RFSKLEAKNLRE 717

RESULT 13
 ALP6_SCHPO STANDARD: PRT: 832 AA.
 ID ALP6_SCHPO
 AC 09USQ2: 09P954;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Spindle pole body component Alp6.
 GN ALP6 OR SPBC902.01C OR SPBC428.20C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RA MEDLINE=20532503; PubMed=11080156;
 RX Vardy L., Toda T.;

RT "The fission yeast gamma-tubulin complex is required in G(1) phase and
 RT is a component of the spindle assembly checkpoint."
 RT EMBL J. 19:6098-6111(2000).
 RN [2]
 RP SEQUENCE OF 1-566 FROM N.A.
 RC STRAIN-972;
 RA Saunders D., Harris D., McDougall R.C., Rajandream M.A., Barrell B.G.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBP databases.
 CC -1- FUNCTION: COMPONENT OF THE GAMMA TUBULE COMPLEX THAT IS REQUIRED
 CC FOR THE REGULATION OF BOTH INTERPHASE MICROTUBULES AND MITOTIC
 CC BIPOLAR SPINDLES.
 CC -1- SUBCELLULAR LOCATION: SPINDLE POLE BODY AND THE MICROTUBULE
 CC ORGANIZING CENTER (MTOC).
 CC -1- SIMILARITY: BELONGS TO THE SPB ALP6/SPC98 FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AB040811; BAA94097.1; ALT_FRAME.
 CC EMBL; AL133306; CAB62095.1; -.
 CC Microtubules; Mitosis.
 KW SEQUENCE 832 AA; 95996 MW; C92771C3DF5C01A CRC64;
 SQ
 Query Match 4.7%; Score 111.5; DB 1; Length 832;
 Best Local Similarity 20.0%; Pred. No. 4.3;
 Matches 101; Conservative 80; Mismatches 194; Indels 129; Gaps 23;

QY 4 SDSVDYTTTLAASEVDSANANAYMNSDYLAVSNDFAERICQVPGK----- 56
 DB 107 SCSRSVSSHLL--DEGISNPINIPSTEVE-----SSNFGQTRYDQVDPENDQITDMD 156
 QY 57 ---SNCASVASVMSRCAQ-----DCLTQSLKPLAKYQPLPLPPYQL 100
 DB 157 EGIENESSISIAHDSRLNSTSTSSVQHTLITLADLLSISYVLOG-----1 204
 QY 101 EAAFLFKESDANPANSTEREFMRFR--GKNHSYFHDL--VFNLKKNYTRDADATDIE- 157
 DB 205 STEYVQFKNELALSKRIPQYQLQMRALSETGLVQELKVFENYDPSVQSISDGNVSK 264
 QY 158 ---NFSRRLY--MATLYITYTNVDEFGASFEFKLSFTTGL-----FQWGIKRAL 203
 DB 265 AFINDOSLALQSLKSVISKELTNFLALIASLDSQIRADASLEKPMYTRRCIMATQVAKL 324
 QY 204 KOIIRSNLPIDIGTEHSVSLQHTSSYKDYMDTOIPALPKFAKRRSLMWVORLLATVAG 263
 DB 325 KLRILSSVANDNMQENKRLIOVSKYVHGDLIOELIS-----DKILTEITG 373
 QY 264 YVDTPWY---KKWYMLKLFNMVNRVEIPTKKFENKEIRPSKALKEKSVSTDTKDFENKI 320
 DB 374 ----PLYEIMENWYK-----GELVDPYQGEF-----VKCKNGSESHD----HQ 409
 QY 321 GGGTV-----DFNKEIRDP--SKALKEKVSNDAKDLFENKIGCGVDFINNEIRDSKA 373
 DB 410 GGGDVVWKKRYPLDKELIPSLSEELVDKIFLIGKSLNFARVCGCDPDM-----AQEHYOK 465
 QY 374 LIRKVSAGADLFEENKIGGTVDFINNEIRDSKALIRKYTEADDFENK-----IGOGT 429
 DB 466 LKAKLSYRPHSLLEYVDKAYTESINHLV-----YLMEEVFLTHDLKAIKTYLLGQG- 519
 QY 430 VDFIN-----KEIRDSKALIR 446
 DB 520 -DFVDLMESSLGNSLDQNPANTLFR 542
 RESULT 14
 ORCL_KL01LA

ID ORCL_KL01LA STANDARD; PRT: 886 AA.
 AC p54788;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-NOV-1997 (rel. 35, Last annotation update)
 DE Origin recognition complex subunit 1.
 GN ORCL.
 OS Kluyveromyces fragilis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxId=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96099401; PubMed=7502077;
 RA Gavin K.A., Hidaka M., Stillman B.;
 RT "Conserved Initiator proteins in eukaryotes."
 RL Science 270:1667-1671(1995).
 CC -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
 CC BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
 CC REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
 CC THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN
 CC ATP-DEPENDENT MANNER.
 CC -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ORC1/CDC6/CDC18 FAMILY.
 CC -1- SIMILARITY: SOME TO YEAST SIR3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U40151; AAC49130.1; -.
 CC InterPro: IPR003959; AAA_subfam.
 CC InterPro: IPR001025; BAH.
 CC Pfam: PR000004; AAA; 1.
 CC Pfam: PR01426; BAH; 1.
 CC SMART: SM00439; BAH; 1.
 CC DNA replication; Nuclear protein; DNA-binding; ATP-binding.
 FT NP_BIND 471 478 ATP (POTENTIAL).
 SQ SEQUENCE 886 AA; 101393 MW; 2DA65ECC291F182F CRC64;
 Query Match 4.7%; Score 110.5; DB 1; Length 886;
 Best Local Similarity 19.5%; Pred. No. 5.4;
 Matches 99; Conservative 82; Mismatches 198; Indels 129; Gaps 23;

QY 5 DSVGDYTTTLAASEVDSANANAYMNSD--MSDYLSAVSNDFAERICQV----- 53
 DB 274 EALSDNESDSEYHESKEEPANANSSDSDEDFEDYQSAEELAYVERAKKVRISIKDIP1 333
 QY 54 -PRGNSCASVASVMS--RCARQDCLLTQSLKRYPLEAKYQ-PLTLPDPYQLBAFLFK 108
 DB 334 SPVKSQTPLOPSAVHSPKRFKNINIVRAKAVTPEKRYKNPKDIPDLND-----IFQR 388
 QY 109 ESDANPANSTEREFMRFRGKNHSYFHDLVFULEENNTVRD-ADATDIEFN--ASRLYIM 166
 DB 389 HNNDDLIALLEERFRVYSAAGKMETTFISKYKOLNSNSKEEIVKADAFNYLPARENEF 448
 QY 167 ATLTYTYTNVDEFGASFEFKLSFTTGLFGWGIKRALKOIIRSNLPIDIT---EHSVSR 223
 DB 449 ASIYLSLSAI--EAGTSTSIYIAGTGV---GTLIVREYVK-----DLMTSADQELPR 499
 QY 224 LOHI-----TSYK-----DYMDTOIPALPKFAKRRSLMWVOR 256
 DB 500 FOYIEINGLKIYKASDSYEVFMOKISGEKLTSGAAMESLEFYFNKYVPATKKRIYVLDE 559
 QY 257 LLATVAGYDTPW-YKKW--YMLKLFNMV--NRVLPTRKFFNK----- 295
 DB 560 LDALVSKQDVWYVNFPMATYVSNAKLIVAVANTLDDLPERHLNKKISSRIGFTIRIMETGY 619

QY 296 ---EIR-----EPKSKALEKYSTDTKLEFNKIGOGTVDFENKIRDP 335
DB 620 THEELRTIINLRKYLNESSFYVDEPCTSSYVISPDSSTI-ETDEEKRPDPSN----- 672
QY 336 SKALEKESYNDAKDLFEKKIGOGTVDFINNEIRDPKSLIRKVSIGADL-----FEKNI 390
DB 673 YKRLKRLINPAIEIASKRIAS---VSGDVRALKVYKVAEYENDYKRLRYERLY 727
QY 391 ---GOGTVDFINNEIRDPKSL 409
DB 728 NSKKTSGNGTNGNELOEVSERIKHTKAL 755
RESULT 15
BACC_BACLI STANDARD: PRT: 6359 AA.
ID BACC_BACLI
AC 068008:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Bactracin synthetase 3 (BA3) [Includes: ATP-dependent isoleucine
adenylase (IleA) (isoleucine activase); ATP-dependent D-phenylalanine
adenylase (D-PheA) (D-phenylalanine activase); ATP-dependent histidine
adenylase (HisA) (histidine activase); ATP-dependent D-aspartate
adenylase (D-AspA) (D-aspartate activase); ATP-dependent asparagine
adenylase (AsnA) (asparagine activase); Aspartate racemase
(EC 5.1.1.13); Phenylalanine racemase [ATP hydrolysing]
(EC 5.1.1.11)].
DE BACC
GN Bacillus licheniformis.
OS Bacillus licheniformis; Bacillus/Clostridium group;
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RX MEDLINE=96089193; PubMed=9427658;
RA Konz D., Klenz A., Schoergerdorfer K., Marahiel M.A.;
RT "The bactracin biosynthetic operon of Bacillus licheniformis ATCC
10716: molecular characterization of three multi-modular peptide
synthetases";
RL Chem. Biol. 4:927-937(1997).
CC -1- FUNCTION: INDUCES PEPTIDE SYNTHESIS, ACTIVATES AND INCORPORATES
FIVE AMINO ACIDS, FORMS A THIAZOLIDINE RING BETWEEN THE FIRST TWO
AMINO ACIDS AND INCORPORATES A D-GLUTAMINE IN THE FOURTH POSITION.
CC -1- CATALYTIC ACTIVITY: L-aspartate -> D-aspartate.
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine -> AMP + diphosphate + D-
phenylalanine
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETHEINES
(POTENTIAL).
CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
ANTIBIOTIC BACTRACIN.
CC -1- SUBUNIT: LARGE MULTIZENYME COMPLEX OF BA1, BA2 AND BA3.
CC -1- DOMAIN: CONSISTS OF FIVE MODULES WITH TWO EPIMERIZATION DOMAINS IN
THE SECOND AND FOURTH MODULES, AND A PUTATIVE C-TERMINAL
THIOESTERASE DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO
THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOALATION, CONDENSATION
(NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
N METHYLATION (OPTIONAL).
CC -1- MISCELLANEOUS: BACTRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
ABUNDANT IS BACTRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT
CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAALOID CONDENSATION
PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-R-PHE-
HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
PHE-9, AND ASP-11).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
FAMILY.
CC -1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF007865; AAC06348.1; -
DR HSSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; DGF4.
DR InterPro: IPR003880; Phosphopant-attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 5.
DR Pfam: PF00668; Condensation; 7.
DR Pfam: PF00550; pp-binding; 5.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 4.
DR PROSITE: PS00455; AMP-BINDING; 5.
DR PROSITE: PS50075; ACP_DOMAIN; 5.
KW Ligase; isomerase; hydrolase; Antibiotic biosynthesis;
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT REPEAT 461 1034 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
FT REPEAT 1517 2064 DOMAIN 2 (D-PHENYLALANINE-ACTIVATING).
FT REPEAT 2999 3570 DOMAIN 3 (HISTIDINE-ACTIVATING).
FT REPEAT 4047 4612 DOMAIN 4 (D-ASPARTIC ACID-ACTIVATING).
FT REPEAT 5549 6129 DOMAIN 5 (ASPARAGINE-ACTIVATING).
FT DOMAIN 1 966 1033 ACYL CARRIER (ACP) 1.
FT DOMAIN 2 1998 2064 ACYL CARRIER (ACP) 2.
FT DOMAIN 3 3502 3569 ACYL CARRIER (ACP) 3.
FT DOMAIN 4 4544 4610 ACYL CARRIER (ACP) 4.
FT DOMAIN 5 6052 6119 ACYL CARRIER (ACP) 5.
FT BINDING 996 996 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2028 2028 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3532 3532 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4574 4574 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 6082 6082 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 6359 AA: 722923 MW: 82A273C546253074 CR64;
Query Match 4.7%; Score 110.5; DB 1; Length 6359;
Best Local Similarity 20.5%; Pred. No. 61;
Matches 117; Conservative 58; Mismatches 173; Indels 223; Gaps 30;
QY 7 VGDVYKTLTLLAASEVSASNAAMVIMNSMDYLSA-----VSDNFAERCSOYPKGSN 58
DB 1897 LGEIENQLKL-DKIDEAAVYARKDDHSDYLCAIYSEKEDMTSTSEISMLEKELP----- 1951
QY 59 CSASVAVMSRCARQDCITLQSLKYLEAKYQPLTPDPYQLEAFLFKESDANPANST 118
DB 1952 -HYMIPATFVR-----LKLPLTSDNDYDKRALPAPBDHVAAGAYEA---PRNDT 1998
QY 119 EKRFEMRRR-----GKNHSYF-----HOLVFLEKN 146
DB 1999 EAKLVIDRDLVAGAGDISHFFFAAGDSIKALQIYSLRLGLKLEMKLFLANPRIKD 2058
QY 147 -----VTRADATDIENFASRYLYMAYTYLKYKTVNDEGASF----- 184
DB 2059 LAKYVKOSORKNANTYITGAHELTPIOKW-----YFA-----NKEELDHNSQSVLEFRK 2109
QY 185 -----FNKL-----SPTTGLFGW-----GIRALK 204
DB 2110 GGFDESCVKKAFNKMIEGHDLRMVIEEKGDFIQYNSFPREDLFDLVYDVRGDRQAE 2169
QY 205 QI-----IRSLPLDIGT-----EHSYSRLQHTSS-----YKDYMDTQI 239
DB 2170 KYVELATSIQKLSSIRKSKLVHLGIFRADEGDHLIIVHLLVVDGVSRLIFEDFETLYS 2229
QY 240 PAL-----PKPAKRFSLMVVORLITAVAGYVDTPMYKKYKMYKLMKNMVRVY 285
DB 2230 QALKGQTLIEIGYKDSYQEFARLKAIVASHTLSKEAEY-----W-----RINAKRV 2277

```

QY 286 -FIPFKFKEIREPSKALKEKYSTD--TKDLFENKIGQTVDFENKEIRDPKSKALKEKY 343
DB 2278 RFLPPLNVKLEDEYENSTLSTLKREATDLRLN-----TNRAYTEIND---ILLTAL 2329
QY 344 SDAKDLE-ENRI-----GQGVDFINN-EI-----RDPKSKALIRKYST 380
DB 2330 LGAGRTIGENLKVMMEHGREDILEGVDITFTIGFTTMYPLVLDAGEKALSOQIKM 2389
QY 381 GAEDL--FENK-IGQTV-----DFINNE 401
DB 2390 VKETLKRIPNKIGYGLTKYMAEDPDFTNDE 2420

RESULT 16
HEKA_IAPUE STANDARD; PRT; 566 AA.
AC P03452; Q83964;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Puerto Rico/8/34).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82013600; PubMed=7278968;
RA Winter G., Fields S., Brownlee G.G.;
RT "Nucleotide sequence of the haemagglutinin gene of a human influenza
  A virus H1 subtype."
RT Nature 292:72-75(1981).
RL [2]
RN SEQUENCE OF 18-343 FROM N.A.
RC STRAIN-SUBTYPE H1.
RX MEDLINE=83129356; PubMed=6186384;
RA Caton A.J., Brownlee G.G., Yewdell J.W., Gerhard W.;
RT "The antigenic structure of the Influenza virus A/PR/8/34
  hemagglutinin (H1 subtype).";
RT Cell 31:417-427(1982).
RL [1]
RN -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V01088; CAA24272.1; -
DB EMBL: J02144; AAA43194.1; -
DR PIR: A04063; HMIV.
DR HSSP: P03437; 5HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN1.
DR PRODOM: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 17
FT CHAIN 18 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 146 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 156 156 A -> E (IN SUBTYPE H1).
FT VARIANT 200 200 A -> P (IN SUBTYPE H1).
FT VARIANT 204 204 D -> E (IN SUBTYPE H1).
FT VARIANT 208 208 I -> L (IN SUBTYPE H1).
FT VARIANT 269 269 R -> M (IN SUBTYPE H1).
FT VARIANT 309 309 F -> Y (IN SUBTYPE H1).
SQ SEQUENCE 566 AA; 63381 MW; E7AE80403547FCE9 CRC64;

Query Match 4.7%; Score 110; DB 1; Length 566;
Best Local Similarity 20.8%; Pred. No. 3.3;
Matches 97; Conservative 67; Mismatches 171; Indels 132; Gaps 24;

QY 23 SAANAVMINSDMSDY-----LSAVSDNFAERISQVPGKS-----NCSASVAYSRCRA 71
DB 101 NSENGICYPEDFDIEBELRQQLSSVS--FER-ELFPKSSMPNHTTGTGYTAACSHAG 157
QY 72 KOD-----CLTQSLKYPLEAKYQPLTLPDYQLEAFLIFRESDA-----NPANST 118
DB 158 KSEFYRNLLMLTEKESYP-----KLKNSYVNRKGKGYLVLMGIHPSNSK 203
QY 119 EKRFMRFRGRKNSHFHDLVFNLKENVTRDADVDIEFASRYLYMATLYKTTTNDV 178
DB 204 DQD---NIYQEN-AYVSVTSNVRFPPEIARERPVROQARMYVWLLRPGDTIIF 259
QY 179 EFGASFF-NKLSPTTGL-FCGMIKRALKOTIRSNLPLDICTEHSVSRLOHTSS--YKDY 234
DB 260 EAAGNLAPRYATAPALSRGREGS-----ITTSNASHHECNTGCQPLGAINSLSPQNT 312
QY 235 MDQIIPALPKFAKRFSLMVV-----QRLATVAGYDTPW--YKKVY----- 274
DB 313 HPVTIGECPRKYVASAKLRMTGLRNIPSIQSRDLFGAIFIGCWGMIDGWYGHQNN 372
QY 275 -----MKLKNFVNRV-----FIPFKFKEIREPSKALKEKYSTD 311
DB 373 EGGSGYAADQKSTQMAINGITNKNVSNVIEKMNIOFAVNGEENK-LEKREINLNKRYVDG 431
QY 312 TKD-----LFEKKIGQTVDFENKEIRDPKSKALKEKYSTDAKDLFENKIGQTV 361
DB 432 FLDIWYMAELLVLEN---ERTLDHDSNVKMLYEYKQLNNAKE-----IGNCFE 483
QY 362 F---INNEIRDPKSKALIRKYSTGAEDL---FENKIGQTVDFINNE 401
DB 484 FYKCKNE-----CMESVRNCTYDYPKYSSESKLNREKVDGKLE 523

RESULT 17
THRT1_SCHPO STANDARD; PRT; 577 AA.
ID THRT1_SCHPO
AC 009684;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nuclear fusion protein tht1.
GN THRT1 OR SPAC13C5.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98106170; PubMed=9442101;
RA Tange Y., Horio T., Shimanuki M., Ding D.O., Hiraoka Y., Niwa O.;
RT "A novel fission yeast gene, tht1+, is required for the fusion of
  nuclear envelopes during karyogamy.";
RT J. Cell Biol. 140:247-258(1998).
RN [2]

```


DB 689 FNNNSDSSAQTLLTKTAATAITSDNEPTREKNNQTKLTTEVEKAKELVEKAKATY 748
 QY 227 ITSSKDYDQTOIPALPKAKRESLWVORL-----ATAVGYDTWYKRYKYLKN 279
 DB 749 SSSQYSEITLLKSSQL-----NDLDDLIILSLDGSIGRTGYANI-----FKMWYKKN 799
 QY 280 FVNRVFIPTKKE-----FNKEIREP-----SKALKEKYSTDKDF 316
 DB 800 STNNEDTSNKEKGEFSPNDLVKQALYIRSMQNLTSKEREGYIDLSVNSTMLQ 859
 QY 317 ENK-----IGQGVDF-----FNKEIRDP-----KALKEKVSND-----KD 349
 DB 860 OKKAQSHTSVNTQLDLAKKAFKELEDPRQDAEYKMYRFLQALMLVKNQAQNYKN 919
 QY 350 LEENKIGQGVDFINNEI---RDPKALIRKYSTGAEIDLENNKIG-----QGT-VD 396
 DB 920 LLOQALPICTRAFSWTVGYDKNPSATVSOQKTSSTSSANENFNFLOPNPYTQSEIN 979
 QY 397 FINNE---IRDPKALIRKYTEADDLFENKI-----GGQTV 431
 DB 980 WENDKOTPIQDPSLSENTYRTDEPNNSVALSKNSGSSD 1022

RESULT 19
 HEMA_IKIE
 ID HEMA_IKIE STANDARD: PRT: 566 AA.
 AC P18875:
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAZ chain].
 GN HA.
 OS Influenza A virus (strain A/Kiev/59/79).
 OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
 OC Influenza virus A and B group: Influenza A viruses;
 OC Influenza A virus.
 NCBI_TaxID=11422;
 RX MEDLINE=86186951; PubMed=3964310;
 RA Beklemishev A.B., Bilnov V.M., Vasilenko S.K., Golovin S.Y.,
 RA Karginov V.A., Mamayev L.V.;
 RT "Primary structure of the full-size DNA copy of the hemagglutinin
 RT gene of Influenza virus A/Kiev/59/79 (H1N1).";
 RL Bioorg. Khim. 12:375-381(1986).
 CC - FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC - SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M38353; AAA43172.1; -.
 DR HSSP: P03437; 2HMG.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein; signal.
 FT SIGNAL 1 16
 FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 346 566 HEMAGGLUTININ HAZ CHAIN.
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 566 AA; 63606 MW; A78A51A4F84D9E74 CRC64.

Query Match 4.6%; Score 108; DB 1; Length 566;
 Best Local Similarity 20.5%; Pred. No. 4.5;
 Matches 95; Conservative 66; Mismatches 151; Indels 152; Gaps 23;

QY 33 DMSDYLSAVS--DNF---AERICQVPGKSNCSASVAYMSKAD-----CLTLQS 80
 DB 117 ELREQLSSVSSEFREFIPKER---SWPK-HNVTREGVASCSSHKGSFYNLLMLEKN 172
 QY 81 LKYPLEAK-----YQPLPLDP-----YQLEAATILPKESDANPANTE-- 119
 DB 173 GSYPLNLSKSYVNKKREVLVLMGVNHPNSNEDQTYRKRENAVSVSSYNRRFPPELA 232
 QY 120 KRFMFRFRGKNHSYFDLVFNLEKNVTRADATDIENFASRYLYMATLYKTYTNVD 179
 DB 233 KRKYVAGQGRINYY-----WTLLEPG-----DTIIFANGNLIA 267
 QY 180 FGASFFNKLSPFTTGLPGWGIKRALKOIRSNLPLDIGTESVSRLOHTSS--YKDYMDT 237
 DB 268 PWYAFALSRGFGSG-----IITSNMSMDCDPCQCPQGAINSLSFQVHVHV 315
 QY 238 QIPALPKFAKRSPLMV-----QRLATVAGYVDTP---YKKWY----- 274
 DB 316 TIGECPRVYVSTKLRLMVTGLRLNIPSIQSRGLFGALFLEGWGTGMDGYGVHNOEG 375
 QY 275 -----MKLNPMVNR-----FIPTKKFFNKIREPSKALKEKYSTDTKD 314
 DB 376 SGYAAQKSTQNAINGITKVNVSIVIEKMTQTFYAVGKEFNK-LEKGMENLNKRYVDGFD 434
 QY 315 -----LFEENKIGQGVDFENKEIRDPKALKEKVSNDKADLENNKIGQGVDF-- 362
 DB 435 IMTYNAELLVLEEN---ERTLDPHDSNNKVLKYEKVSQKLNNAKE-----IGNGCFEYH 486
 QY 363 -INNEIRDPKALIRKYSTGAEIDL---FENKIGQGVDFINNE 401
 DB 487 KCNNE-----CMESVKNQTYDPRYSESKLNREKIDGVKLE 523

RESULT 20
 YB47_METJA
 ID YB47_METJA STANDARD: PRT: 462 AA.
 AC Q58547;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1147.
 GN MJ1147.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 NCBI_TaxID=2190;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,
 RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Huret M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67557; AAB99156.1; -.
DR TIGR: Mj1147; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 433 453 POTENTIAL.
SQ SEQUENCE 462 AA: 54433 MW: 91608862971776 CRC64;

Query Match 4.6%; Score 107.5; DB 1; Length 462;
Best Local Similarity 22.7%; Pred. No. 3.8;
Matches 44; Conservative 30; Mismatches 75; Indels 45; Gaps 7;

QY 130 KNHSFHLVNLKENTROAD-----ATDENFASRYLMATLYXTYTVNDEFGASF 184
DB 135 KSHVEFDISFNLSTFDCRRDDVSFKKIDKNEKGLAFLKTFPLKKHTTENFOLS- 253
QY 185 FNKLSFTTGLFGWGIKRAKQIIRSNLPDIGTE---HSVSRLOHITSSYKDYMDTOIP 240
DB 254 --KISF-----LKTIDVREVLLCDVYKKEILSHKIRIKEDSGNKKDKLENLK 299
QY 241 ALPFAKPSLMAVORLLATVAGYVDTPWYKKWYKLNPMV---NRFLEPTKFFPK 296
DB 300 ELGLSLSYR-----IIDQFNYSVLAEYRNLRSISINNNTYIASMLYME 345
QY 297 ---IREPSKALKER 307
DB 346 MELIKESNGRPEK 359

RESULT 21
HEMA_IJAH3
ID HEMA_IJAH3 STANDARD: PRT: 550 AA.
AC P11134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/136/82).
OS Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT in China.";
RL Virology 162:160-166(1988).
RU
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19056; AAA43211.1; ALT_TERM.
DR PIR: A29971; HMV52.
DR HSSP; P03437; ZHMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NOW TERM
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA: 61580 MW: 9916088629724F2 CRC64;

Query Match 4.5%; Score 107; DB 1; Length 550;
Best Local Similarity 18.4%; Pred. No. 5.1;
Matches 95; Conservative 63; Mismatches 164; Indels 194; Gaps 23;

QY 5 DSVGVYTKTLAASESVDSANAYMNSDMSDYLS-----AVSD 43
DB 77 DVFGQETMDLEVERSNARSCYPP---DVPDYASLSRLVASSGTFLEGTGFTVGTQ 132
QY 44 NFAERICQOVKSGMSASVAYMSRCAKQDCPLQSLKYPLEAKYQPLTLPDPYQLDLA 103
DB 133 NGSSACKRGP-----ASGFESRL--NWLTKSGSTYPV---LNVTPMNDNPKL 177
QY 104 FL-----LFKESDANPANSTK-----RFWRRFRGKNHSYFHL 138
DB 178 YIMGVHHPSTNOEQTNLVQASGRVTVSTRSQQTIPNIGSRPVRQSGRISITW--- 234
QY 139 VFNLEKNTVDADATDIENFASRYLYMATLYKTYTVNDEFGASFENKLSFTTGLFGWG 198
DB 235 ---TIVKPGDVLVINSNGN--LIAPRGYFKMRT-----GKS----- 265
QY 199 IKRAKQIIRSNLPDITENHSVSRLOHTSS--VKQDMYDQIPALPKFAKFSIMV--- 253
DB 266 -----SIRSDAPIDTCVSECTIPNGSIPNDKPPQVNNKITTYGACPXYVQNSLKLATG 319
QY 254 -----VORLATVAGYVDTPW---YKKWY-----MKLK-----NF 280
DB 320 MRNVPEKQTRGLFGAIAFIENGWEGMIDGWYGFRRHNSSEGTGQAADLKSQAATDQING 379
QY 281 MYNRVFIPTKKFFN---KEIRPSKALK--EKVSTDTKDLFENKIGQTVDFENKEIRDP 335
DB 380 KLNRIEKTNEKHQIEKFESEVEGRIDLEKYVEDTK-----IDLMS----- 422
QY 336 SKALKKESVNDAKDLFEKKIGQTVDFINNEIRDSKALLIRKYSNGADLFEKKIGQTV 395
DB 423 -----YNADVLALEN---OHTIDLDSSENNKLFETKTRQLRENAEDM-----GNGCF 467
QY 396 DFINNEIRDSKALLIRKYVTEADLFEKKIGQTV 431
DB 468 -----KIYHKCDNACISIRNGTID 487

RESULT 22
RRPL_DUGBV
ID RRPL_DUGBV STANDARD: PRT: 4036 AA.
AC Q66431.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (L protein).
GN L.
OS Dugbe virus.

OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Nairovirus.
OX NCBI_TaxID=11595;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE ARD 44313;
RX MEDLINE=96332515; PubMed=8760425;
RA Marriott A.C., Nuttall P.A.;
RT "Large RNA segment of Dugbe nairovirus encodes the putative RNA
polymerase."
RL J. Gen. Virol. 77:1775-1780(1996).
CC -i- CARALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
[RNA](N).
CC -i- SIMILARITY: CONTAINS 1 OTU DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: U15018; AAB18834.1; -
DR InterPro: IPR003323; OTU.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF02338; OTU.1.
DR PROSITE: PS50802; OTU.1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2.1; 1.
DR Transferrase: Nucleotidyltransferase; RNA-directed RNA polymerase.
KW DOMAIN 29 158 OTU.
SQ SEQUENCE 4036 AA; 459382 MW; E2EDF0B4358E31BD CRC64;

Query Match 4.5%; Score 106.5; DB 1; Length 4036;
Best Local Similarity 20.4%; Pred. No. 64;
Matches 75; Conservative 64; Mismatches 141; Indels 87; Gaps 17;

OY 148 TRDADATDI--ENFASRYL-----YMATLYKTYTNVDEFGASFNKLSFTTGLG 196
DB 271 TRESEATISLKLQGHKLLTRDHIMETARSKLYTRDLIDLDIGGLRSSFP---G 326
OY 197 WGIKRAKQIIRSNPLDIGN-----EHSVRLQIHITSSYKDYMDTOIIPALP 243
DB 327 LGHERYI-QLHSELVLDLVTVLAVLSTFLGSGNNKKNKQFITNC--LMTKLSGR 382
OY 244 KPAKRSLMNVQRLAT---VAGYVTPMYTKKYMKLKFM-----VNRVFTPKKFFNK 295
DB 383 VF-KALSKLTGQMLYTPRAVSIVSOELYKMLKVKNNLECMGPISMLALNLFNDM 441
OY 296 EIREPSKALKEKYSTDTKDLFENKIGGYDFENKEIRD-----PSKALKEKVSNDAK 348
DB 442 QLODYILEMSEKIDNSD-----VEYTHREISDLHTLVERLSKLQKSDYNEIK 491
OY 349 DLF-----ENKIGQGVDFINNEIRDPKSKALIRKVSSTAEDLFENKIGQGVDFIN 399
DB 492 LMFKEEVLTKRSQSYGNFEEFLINDYFK--KKDIMKPVSTGSKASTGNIG-NVLSYAH 548
OY 400 N-----EIRDPKAL-----IRKVTYTDADLFENKIGQGVDFINKEIRDPKSKALIRK 447
DB 549 NYLSKESLRMSDEVTOILLIEIRKLKLOGDLSEPPVA-----IICDKLEDFRKLRE 603
OY 448 VSTEADN 454
DB 604 LPRECS 610

DE Hemagglutinin HA2 chain] (Fragment).
GN HA
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks".
RL Virology 159:109-119(1987).
CC -i- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -i- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -i- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: M16742; AAA43148.1; -
DR PIR: F27813; HMIY98.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Hemagglutinin; glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328
FT CARBOHYD 330 550
FT CARBOHYD 22 22
FT CARBOHYD 38 38
FT CARBOHYD 165 165
FT CARBOHYD 285 285
FT CARBOHYD 483 483
FT CONFLICT 8 8
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 4.5%; Score 106; DB 1; Length 550;
Best Local Similarity 18.3%; Pred. No. 5.9;
Matches 92; Conservative 65; Mismatches 153; Indels 192; Gaps 23;

OY 19 ESDVSAANAMINMSDMSDIYS-----AVSDFARISQVPPKGS 57
DB 89 ERSNAFSNCPY--DVPDVYSLRSLVASSGTLFEITGFTWGTQVTONGSGNACKRGPB- 144
OY 58 NCSASVAVMRCADKDCLOLSLKYPLFAKYOPLTLPDYOLEAFI----- 105
DB 145 -----SGFSSRL--NMLTKSGSTYV---LNTVMPNNDNDKLYIYGVHHPSTNOEQ 191
OY 106 --LFKESDANPANSTK-----RFWRFRRGNKNSYFHDVLVFNLEKNVTRDAD 152
DB 192 TNLVVOASGCVTVSTRSOOTIIPNIGSRPVRVQSGRISYIY-----TVVAPGD 241
OY 153 ATDIENPASRYLYMATLYYTYTYTNVDEFGASFNKLSFTTGLGCGWGIKRAKQIIRSNLP 212
DB 242 VLVINSNGN--LIAPKGYFMRT-----GKS-----SIMSDAP 273
OY 213 LDIGTEHSVRLQIHITSS--YKDYMDTOIIPALPKFAKFSILAV-----VORLLA 259
DB 274 IDICISECITPNSGIPRDKRFQVNVNKKTYGACPKRYKQNTLKLATGMRNVPEKOTGLG 333
OY 260 TVAGYVDFPW---YKKWY-----MKLK-----NPNVNVFPIPTKKFFN 294
DB 334 AIAGFIENGEWGMIDGWYGRHONSEGTGAADLKSFOAIDQINGKLNHVIETKEKFFH 393

```

OY 235 ---KEIRPEKALK--EKVSTDTKDLFENKIGQTVDFPNKEIRDPKALKKEKVSNDAKD 349
DB 334 QIEKEFESEVEGRIOLEKEVEDTK-----IDLMS-----YNADVLV 429
OY 350 LEENKIGQTVDFINNEIRDPKALKIRKYSTGAEDLFENKIGQTVDFINNEIRDPKALK 409
DB 430 ALEN---QHTIDLTSEMKKLFERTRRQLRENAEDM-----GNGCF----- 467
OY 410 IRKVTYEAADLFENKIGQTVDF 431
DB 468 --KIYKCDNVCIESIRNGTYD 487

RESULT 24
HEMA_IADH7 STANDARD: PRT; 550 AA.
ID HEMA_IADH7 084018: 089470:
AC P12588; 084018: 089470:
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M16743; AAA43149.1; -.
DR PIR: G27813; HMIY15.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 550 AA; 61761 MW; 6EE81793281D53EB CMC64;

Query Match 4.5%; Score 106; DB 1; Length 550;
Best Local Similarity 18.3%; Pred. No. 5.9;
Matches 92; Conservative 65; Mismatches 153; Indels 192; Gaps 23;
OY 19 ESDVSAANAYMINDMSDLS-----AVSDNFAERICSGVPGKS 57
DB 89 ERSNAFSCYPT--DVPTASLRLSVASSGTLFTTGGTWTGVTONGSGNACKRGPN-- 144

```

```

OY 58 NCSASVAYSBCAKODCLTIQSLKYPLEAKYQPLTLPDYOLEAFI----- 105
DB 145 -----SCFESRL-----NMLTSGSTYPV-----LNTWPMNDNFDKLY IGVHHPSTNQ 191
OY 106 ---LFKESDANPANSTK-----RWMRFRRGKHSFPHDLYVFLKKNYTRQD 152
DB 192 TMLYVOAGSRVTVSTRSSOOTIIPNIGSRPWVGOSGRISIYW-----TVKPGD 241
OY 153 ATDIENFASRYLYMATLYKTYTNVDEFGASFENKLSFTTGLFGWIKRALQIIRSNLP 212
DB 242 VLVINSGN--LIAPRGYKMT-----GKS-----SIMRSDAP 273
OY 213 LDIGTEHVSRLQHTSS--YKDYMDTOIPALPKFAKPSLMV-----VORLLA 259
DB 274 IDTCISECTTPNGSIPNDKPFQNVNKITYGACPKYKQMTLKLATGMNVPEKOTRGLFG 333
OY 260 TVAGYVDTPW---YKKWY-----MKLK-----NPMNVRVYLPKPKFFN 294
DB 334 ATAGFLENMCEGMIDWYGFRRHONSEGTOAADLSTQAIIDQINKLNRVLEKTEKPH 393
OY 295 ---KEIRPEKALK--EKVSTDTKDLFENKIGQTVDFPNKEIRDPKALKKEKVSNDAKD 349
DB 394 QIEKEFESEVEGRIOLEKEVEDTK-----IDLMS-----YNADVLV 429
OY 350 LEENKIGQTVDFINNEIRDPKALKIRKYSTGAEDLFENKIGQTVDFINNEIRDPKALK 409
DB 430 ALEN---QHTIDLTSEMKKLFERTRRQLRENAEDM-----GNGCF----- 467
OY 410 IRKVTYEAADLFENKIGQTVDF 431
DB 468 --KIYKCDNVCIESIRNGTYD 487

RESULT 25
YH08_YEAST STANDARD: PRT; 624 AA.
ID YH08_YEAST 054074:
AC P54074;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 71.4 kDa protein in ILV2-ABE17 intergenic region.
GN YMR119W OR YMR718.18 OR YMR564.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-431 FROM N.A.
RA STRAIN-S288C / AB972;
RC Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 345-624 FROM N.A.
RA STRAIN-S288C / AB972;
RC Lye G., Churche C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO YEAST YNL008C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: 249702; CAA89757.1; -.
DR EMBL: 249773; CAA89268.1; -.
DR SGD: 50004725; YMR119W.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 32 52 POTENTIAL.

```

FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 FT TRANSMEM 274 294 POTENTIAL.
 SQ SEQUENCE 624 AA: 71443 MW: E2F5149F89A09655 CRC64;

Query Match 4.5%; Score 106; DB 1; Length 624;
 Best Local Similarity 25.6%; Pred. No. 6.9;
 Matches 68; Conservative 43; Mismatches 87; Indels 66; Gaps 17;

QY 171 YKTYTNVDEFGA---SFFNKLSTTGLFGWGIRKALKQIIRSNLPDICTEHSV-SRLQ 226
 DB 321 YLNTGGEEDFSAMATKFAALLCSGTLMKGTIRREPAI---NIPNNVEKFFISGLY 377
 QY 227 TTSSTKDYDTQIPALPK-----FAKRESLAVVQRLATVAGVDTPTWKYKKWIKRNP 281
 DB 378 LSKPYKE--NTSI-SFFKNSSILKORFIMPEKSI-----WMKK---L 417
 QY 282 VNRVFIPTKKFFNKET--REPSKALKEVSTDKDLFENKIGQ-----GTVDFFNK 330
 DB 418 VQVTFEGFDNKDEDIPNDPSPKMLK---ITKTNSL--NNSAGHKEDIETELNTSDEYS 473
 QY 331 EIRDSPSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPKALIRKYSTGAEDLEPKRI 390
 DB 474 EDERPSE-VESLGDSEENLEEDSL-----IFNETRD--ALL-----DLFSSD 514
 QY 391 GGGTVDFINNEIRDPKALIRKYVTE 416
 DB 515 NEVHTDY--NWIMSTSRILQOKRLSD 538

RESULT 26
 HEMA_IADHK STANDARD: PRT: 550 AA.
 AC P43257;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC Influenza A virus.
 OX NCBI_TaxID=11364;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 of avian H3 influenza viruses to pigs in southern China, where the
 A/Hong Kong/68 (H3N2) strain emerged.";
 RT J. Gen. Virol. 72:2007-2010(1991).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC or send an email to license@sib-sib.ch).
 CC EMBL: D00929; BAA00769.1; -
 DR HSP; P03437; SHMG.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1 1
 FT CHAIN 1 328
 FT CHAIN 330 550
 FT CARBOHYD 8 550
 FT CARBOHYD 8 550
 FT CARBOHYD 22 22
 FT CARBOHYD 38 38
 FT CARBOHYD 165 165
 FT CARBOHYD 285 285
 FT CARBOHYD 483 483
 SQ SEQUENCE 550 AA: 61549 MW: 864639829FEI8A9 CRC64;

Query Match 4.5%; Score 105.5; DB 1; Length 550;
 Best Local Similarity 18.0%; Pred. No. 6.4;
 Matches 88; Conservative 69; Mismatches 154; Indels 177; Gaps 22;

QY 12 KTLTAAESVDSANAYMINSMDYLSAVSDNFARICSOVPGKSNCSASVAYSRCA 71
 DB 109 RSLVASSGTLEFTTBFT-----WTGVTONGSSACKRGF-----ASGFSRL- 151
 QY 72 KODCLTLQSLKYPLEAKYPLTPDYQLEAFI-----LKFESDANPANS 117
 DB 152 --NMLTKSGSTYFV---LNVTPNNDNFDKLYIMGVHHPSTNOEQTNLVQASGRVTS 205
 QY 118 TEK-----REWMRFRGKNHSYFHDLYFNLEKNVTRDADATDIEFNASRYLYM 166
 DB 206 TRRSQGTIIPNIGSRPWRVQSGRISYV-----TIYKGVLYVNSGN--LIA 253
 QY 167 ATLYKYTYNVDEFGASFENKLSFTTGLFGWGIRKALKQIIRSNLPDICTEHSVRLQ 226
 DB 254 PRGYFKMRT-----GKS-----SIMRSDAIPIDCVSCLIPNPS 287
 QY 227 TTS--YKDYMDTQIPALPKFAKRESLAV-----VQRLATVAGYDTPW--Y 270
 DB 288 IPNDKPFQVNVKITTYGACRKYVQNSLKLATGRNVPEKOTRLFAIAGFIENGEGMI 347
 QY 271 KKYV-----MKLK-----NEMVNVFIPTKKFFN---KEIRPSKALK 305
 DB 348 DGYGFRHNSSEGTGAADLKSTQAAIDQINGLNRYIKTKNEFKQIEKSEVSGRIQ 407
 QY 306 --EKVSTDKDLFENKIGQGTVDFFNKETIRDPKALKEKVSNDAKDLFENKIGQGTVD 363
 DB 408 DLEKYVEDRK-----IDLMS-----YNADVLALEN--QHTIDLT 440
 QY 364 NNEIRDPKALIRKYSTGAEDLEFENKIGQGTVDFINNEIRDPKALIRKYVTEADLFEN 423
 DB 441 DSEMNKLFETKROLRENAEDM-----GNCF-----KIYHCKDNACIE 479
 QY 424 KIGQGTVD 431
 DB 480 STRNGTYD 487

RESULT 27
 HEMA_IAENT STANDARD: PRT: 566 AA.
 AC P03440;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/England/321/77).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC Influenza A virus.
 OX NCBI_TaxID=11378;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83110955; PubMed=6822816;
 RA Hauptmann R., Clarke L.D., Mountford R.C., Bachmayer H., Almond J.W.;
 RT "Nucleotide sequence of the haemagglutinin gene of influenza virus

RT A/England/321/77.":
RL J. Gen. Virol. 64:215-220(1983).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05907; CAZ9337.1; -.
DR PIR: A04052; HMY6.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR PRODOM: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63608 MW; FA5B886F4B8C888 CRC64;

Query Match 4.5%; Score 105.5; DB 1; Length 566;
Best Local Similarity 19.1%; Pred. No. 6.6;
Matches 103; Conservative 73; Mismatches 185; Indels 177; Gaps 27;

OY 11 TKTLLAASEVSASANAAMVINSMSDYLSA---VSDNFAERICS---OYPKSNCASASV 64
DB 26 TATLCLAHAVNGTLVITTDQIEVNTATLVOSSSTRGICDSPHRLDKNCTL-ID 84
OY 65 AVMSCANODCTLTLOSLKYLEAKYO-----PLLPDPYOL-----EAAFILEKESDA 112
DB 85 ALL---GDPHCGFQNEKMDLFEVRSKAFSNCYPYDVPDYLASRLVASSGLEFINEGF 141
OY 113 NPANSTEFKFMFRGRGNHSHFDLVFNLEK-----NVRDADATDIENASRL 164
DB 142 NMTGVATONGSYACKRGPDNSFFSRL--NMLYKSESTYPLVLTVMNPN---DNFDKLYI 195
OY 165 Y-----MATLYYKT-----YTNDDEFGA-SFFKKLSFTTGLFGM---- 197
DB 196 MGVHHPSTDKETKLYVASGRVYSTRSQOTIIPNGSRWVAGLSRISIT-YTIYK 254
OY 198 -----GIRALKOIIRSNLPDLDIGTEHSVRLQHTSS--YKDYM 235
DB 255 PGDILLINSNGNLIAPRGYFKIRTKSSIMRSDAPIGTCSSCIPNCSIPNDKDFQNVN 314
OY 236 DTQIPALPKFAKRFSLM-----VQRLLATVAAGVDTPM--YKKKY----- 274
DB 315 KITYGACPRYVAKQNTLKLATGKRNVPKQTRGIFGALIGFLENGEMGIDWGFRRHONS 374
OY 275 -----MKLK-----NEAVNRVFIPTKKFEN---KEIREPKALK--EKVSTDFK- 313
DB 375 EGTGAADLKSTQALIDQINGKLNIVIEKTNKEFKHOIEKESFEVSGRLODELEYEDTKI 434
OY 314 DLFEKKIGOGVDFPNKEIRDSKALKEKVSNDADKLFENKIGOGTVDFINNEIRDSPSKA 373
DB 435 DLMS-----YNAEL--LVALE-----NQHTIDLTDEMKMLEEK 466
OY 374 LIRKYSTGAEDLFEKKIGOGTVDFINNEIRDSKALIRKYTEADDLFEKNKIGOGTV 431

DB 467 TRROLENNEDM-----GNGCF-----KIYKCDNACIGSRNGTYD 503
RESULT 28
ID Y075_MYCPN STANDARD; PRT: 1030 AA.
AC P75556;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG075 homolog (G07_Orfi1030).
GN MPN213 OR MP618.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000059; AAB96266.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 29 POTENTIAL.
SQ SEQUENCE 1030 AA; 116017 MW; DA4E48035973E07C CRC64;

Query Match 4.5%; Score 105.5; DB 1; Length 1030;
Best Local Similarity 21.0%; Pred. No. 14;
Matches 92; Conservative 71; Mismatches 171; Indels 105; Gaps 21;

OY 78 LOSLKYPLEAKYOP-----LTLPD-----PYOLEAFILEKE-----SD 111
DB 51 LMSVKLP---KAQPAATRTIVENGDKLVNKKSSPOQLFLAKNALKDLOGEFDFKLS 107
OY 112 AN--PANSTEFKFMFRGRGNHSHFDLVFNLEKNTVRDADATDIENFASRYLYMATL 169
DB 108 AKAPFALTAADLOEWVDOQLFNPDSFFDL--SAPRSNFTLSSDKRASIDFIR----- 158
OY 170 YKTYTNDVEFGASFENKLSFTTGLFGMGIRALKOIIRSNLPDLDIGTEHSVRLQHTS 229
DB 159 ----FTN-----FTESVQLKLPREGVSVYDSSKSPFYVYVNASQKLLVPL 201
OY 230 SYKD-----YMDQIPALPKFAKRFSLMVYQRLLATVAAGVDTPMYKKMYKLNFVN 283
DB 202 SLPDYTLGLNMYFHDITLNGKVNKFSFNPFTNINLA-----FSNVYGVGVDFEAO 253
OY 284 RVFIPTKKFENKEIREPSKALKEKYSTDKDLFE-----NKIGOGTVDFPNKEIRDSK 337
DB 254 KNLVKGKRYLNVHVK--MEDVKRDVANNTKNOFDIAKIIAELMGKALKEFGNQDGPPLS 311
OY 338 ALK--EKVSNDADKLFENKIGOGTVDFINNEIRDSKALIRKYSTGAEDLFEKNKIGOGTV 395
DB 312 FLKWDKYEKDEKLF-NLVRPGLGKFPYKDLIOSSQA--EKKITVYKLIIFPNK---KTI 365
OY 366 DEINNEIRDP-----SKALIRKYV---TEADDLFEKNKIGOGTVDFINKEIRDSKA----- 443
DB 366 LNLKELSIPELNSSLGLVDFDGTDSOGYERL--QSFKDLIVPAVKTNKEKTAALSP 423
OY 444 LIRKYSTGAEDLFEKNKIGOGTVDFINNEIRDSKALIRKYTEADDLFEKNKIGOGTV 431


```

Db 423 ---IYKLRRIESTEDKDYVETLNDKIESLEKQOLDKQNLNRELRSISKIDETTORYL 479
Oy 125 RPRRCNHSYFDLVNLEKVNTRDADATDIENFASRYLYAKTYTNVDERGASF 184
Db 480 DAK-----DIELKQSVSIRKQETLAKMKAQLELSAIIQOT-----515
Oy 185 FFKLSFTTGLFGWGIKRALKQIIRSNLPIDIGTSHS---VSRLOHTTSYKDYMDTOJP 240
Db 516 VVELTAT-----KNELSQVATTNQSLFAENEESKVLLEGTIARIDSFYQIIMQIEKP 567
Oy 241 ALPKRAKRSRLMNVQRLATVAGYDTPMYKKMKMLKFMNRR-----284
Db 568 DYVPIISK-----POLTNQESIT-OTDIKQMDLQKSLNADYANLQSVSEISDLKS 619
Oy 285 --VFIPTKKFKFNKEIRPSKALKKESVTDTKDLFENKIGQVDFENKIRPSKALKRK 342
Db 620 QIISIVPRIVIRILKENKAKVENNNA---ELLESVAITSVLSAL---VOQYERSEKON 673
Oy 343 VS-----NDADLFENKIGQVDF--INNEI-RDPSKALKRKVSTGAEDLFENKI 390
Db 674 VKLRQEFELKLDLQRLLE---ONQTFESISEFISRPD--AFNNLN---DERFONLR 724
Oy 391 GGGTVDFTINNEIRDPSKALKIRKYTRADDLFENKIGQVDFINKEIRPSKALKRKVST 450
Db 725 QO--YDEMSSKYSALETTKIKEMESTADQAVKSEMSK-----LNTQL-DELNSLFVKYNR 776
Oy 451 EADNMLE 457
Db 777 KADIFE 783

RESULT 31
AC NFM_CHICK STANDARD: PRT; 857 AA.
AC P16053;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NFM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174973; PubMed=2106668;
RA Zopf D., Dineva B., Betz H., Gundelfinger E.D., Betz H.;
RT "Isolation of the chicken middle-molecular weight neurofilament
RT (NF-M) gene and characterization of its promoter.";
RL Nucleic Acids Res. 18:521-529(1990).
RN [2]
RP SEQUENCE OF 259-857 FROM N.A.
RX MEDLINE=88112814; PubMed=3123320;
RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT "Identification of gene products expressed in the developing chick
RT visual system: characterization of a middle-molecular-weight
RT neurofilament cDNA.";
RL Genes Dev. 1:699-708(1987).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PFM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PFM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X17102; CAA34958.1; -
DR EMBL: X05558; CAA29073.1; -
DR PIR: A27040; A27040.
DR PIR: S08061; S08061.
DR PIR: S15762; S15762.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE, PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurex; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL_1A.
FT DOMAIN 131 143 LINKER_1.
FT DOMAIN 144 242 COIL_1B.
FT DOMAIN 243 259 LINKER_12.
FT DOMAIN 260 281 COIL_2A.
FT DOMAIN 282 285 LINKER_2.
FT DOMAIN 286 406 COIL_2B.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SQ SEQUENCE 857 AA; 95704 MW; 4E2E0FC6AC64778B CRC64;

Query Match 4.5%; Score 105; DB 1; Length 857;
Best Local Similarity 18.1%; Pred. No. 12;
Matches 94; Conservative 78; Mismatches 183; Indels 164; Gaps 20;

Oy 2 APSDSVGDVTKTLAASESDV-----SAANAAYINSDMSDYLSAVSDNFA---ERIC 50
Db 58 APRTAG---STVLSAESELDVSSSLNGAAMELKRSNENKQGLGDRFAGYEKV- 113
Oy 51 SOVPKGSNCASVAYMSKACODCLTLOSILKYPLEAKQPLTLPDPYOLEAFILFKES 110
Db 114 -----HYLEDQNKIEBALEALRQKNGRAQ---LGDVDEQ-----147
Oy 111 DANPANSTKRFMRPRRCNHSYFDLVNLEKVNTRDAD--ATDIENFASRYLYMAT 168
Db 148 -----LRELKALDEOVSH-----KAOIQDSEHIEDIQRLERFEDEAR 188
Oy 169 LYYKTYTNV-----DEFGASFNKLSTFTGLFGWGIKRALKQIIRSNL 211
Db 189 LDDEFTATIALRKEMERASLMRAELDKKVQSLQDEAVLFRGNHEEVALLAQL-----243
Oy 212 PLDIGTSHSVSRLOHTTSYKDYMDTOI-PALPKRAKRSRLMNVQRLATVAGYDTPMY 270
Db 244 -----OASHATYERKDYLTALKEIRAQLECSQDNHM-----HOAEWF 286
Oy 271 KKWYMLKKNFMNVRVFIPPKFENKIRPSKALKKESVTDTKDLFENKIG---OGTVD 326
Db 287 KCRYAKLTFAA-----EONKFAIRSAKEELIEYRQLOSKSIELESVRGTRKE 333
Oy 327 FENKIRPSKALKKESVNDADLFENKIGQVDFINNEIR-----368
Db 334 SLERQSD-----IEERHNNDLTY-----QDTIQLNELGELGTIKEMARHLREYQDLLN 383
Oy 369 -----DPSKALKRKVSTGAEDLFENKIGQVDFINNEIRDPSKAL---IRKYTRAD 419
Db 384 VKMALDIEIAAARKLLEGEETRFSAFSGSITGPIFTN--RQPSVYIATYKIQKTYIEPK 441
Oy 420 L-----FENKIGQVDFINKEIRDPSKALKRKVSTEA 452

```

DB 442 LKVOHKVEEIEETKEVEDEKSEMEDALSAIAEEMAKA 480

RESULT 32

YF05_METUA STANDARD; PRT; 778 AA.

AC 058900;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative ATP-dependent RNA helicase MJ1505.

GN MJ1505.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

OC Methanococcus.

OX NCBI_TaxId=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RT Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii."

RL Science 273:1058-1073(1996).

CC -! SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U67591; AAB99518.1; -

DR TIGR: MJ1505; -

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002464; DEAH_ATP_helicase.

DR InterPro: IPR004016; ERCC4.

DR InterPro: IPR000445; HHH.

DR InterPro: IPR003583; HHH_1.

DR InterPro: IPR001650; Helicase_C.

DR Pfam: PF00270; DEAD; 1.

DR Pfam: PF02732; ERCC4; 1.

DR Pfam: PF00271; helicase_C; 1.

DR Pfam: PF00633; HHH; 1.

DR SMART: SM00487; DEXDC; 1.

DR SMART: SM00490; HELIC_C; 1.

DR SMART: SM00278; HHH1; 1.

DR PROSITE: PS00690; DEAH_ATP_HELICASE; FALSE_NEG.

DR Hypothetical protein; ATP-binding; RNA-binding; Helicase;

KW Complete proteome.

KM NP_BIND 35 42 ATP (POTENTIAL).

FT SITE 137 140 DEAD BOX.

FT DOMAIN 532 535 POLY-GLU.

SQ SEQUENCE 778 AA; 88882 MW; DCBIF2208013368 CRC64;

Query Match 4.4%; Score 104.5; DB 1; Length 778;

Best Local Similarity 18.5%; Pred. No. 11;

Matches 103; Conservative 91; Mismatches 163; Indels 199; Gaps 28;

OY 5 DSVGDVTKTLAASEVDSANAMVINSDSYLSVSNFAERICSQVKGSCNSASVS 64

DB 242 NSIADVTKELE-----LNKKLSYDEEVKVELI-KVCEALKLMHARELLE 288

OY 65 A-----YMSRCADCLTLOSLEYEAKYQPLTPDPYOLEAFLFKESDANPAN 116

DB 289 SGGKSVFLANIINKLSMQRTSAKSI-----VDEKREAVNLLMKSVDEP-- 334

OY 117 STEKRFMRFRCKNSYFDLVNLEKAVTADADATDIENFASRLVATLYKTYTN 176

DB 335 -----KLGK-----VVDVKNILEKN-KD-----ERLIIPAQ-YRDVER 367

OY 177 VDFEGASFNNKLFPTTGLFGWGIKRALQIIRSLPLDICTEHSVSLQHTSSYKD-- 233

DB 368 I-----VNLTLQNGIKAIRITGQANKEGK 392

OY 234 -YMDTQIPALPKFKRESLNV-----VORLLAVAG 263

DB 393 MSQKEQLEALERFKKBESVLVTSVSEGDIPSVMIIIEPVPSEIRIQGRAMRG 452

OY 264 -----YDTPMKKMYK-LKN-FMVNRVFIPTKFKFNK---EIRE 299

DB 453 EGGKVVYLIAKGTADENAYRSALYKEREMRLKMKCYLLINKRL--OKKFEESKEBIKE 510

OY 300 PSKALKK-----VSTDKDLFENKIGGV-DFPNK-EIRDPKALKKESNDK-- 348

DB 511 ETEELKEKTESYAVKEETKEDEKTKPVTIDFTKOLEVERKSSEEDKIQETKIP 570

OY 349 -----DLFENKIGGVDFINNEIRD-----PSKALKRYSTGAEDLFENKIGQ 392

DB 571 KKPKTIIVDRKMKMLHNANIELKTEVGVDVLSDRVVERK--ADEPVNSIIDK 628

OY 393 GTVDFINNEIRDPKALKIRVYTEADDLFENKIGGV----DF-----INKEIRDPK 442

DB 629 RLFSQLEKN-LKVEKPLIVEGENFSRLHENAL-KGALISIIDFGPIIFTKNAEETAD 686

OY 443 ALIRKSTENDNLEK 458

DB 687 LLIR-LAEKQIRK 700

RESULT 33

HEMA_IADH1 STANDARD; PRT; 550 AA.

ID HEMA_IADH1

AC P12582; Q84021; Q84022;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/5/77).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A virus.

OX NCBI_TaxId=11357;

RN SEQUENCE FROM N.A.

RP MEDLINE=87265458; PubMed=2440178;

RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild

RT ducks."

RL Virology 159:109-119(1987).

CC -! FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -! SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -! SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

EMBL: M16737; AAA43143.1; -
 DR PIR: A27813; HMI177.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR Prodom: PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 8 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61705 MW; 7E7ACEF716FC969A CRC64;

Query Match 4.4%; Score 104; DB 1; Length 550;
 Best Local Similarity 18.1%; Pred. No. 8;
 Matches 91; Conservative 65; Mismatches 154; Indels 192; Gaps 23;

19 ESDSAAANAYMINSMDYLS-----AVSDNFAERICSOVPGKS 57
 DB ESNMNFNCYPR--DVPDVAISRLVASSGTLFTEGTTWGTGONGSSNACKRGP--- 143
 QY 58 NCSASVSAVMSRCANQDCITLQSLKYPLEAKYQPLTLPDPOLEAFT----- 105
 DB 144 -----ASGFFSRLL--NMLTKSGSTYPV---LNVTPMNDNFDKLYINGVHHPSTQEQ 191
 QY 106 --LFKESDANPANSREK-----RPMRFRGKNHSEFHDLYVFLLEKNYTRDAD 152
 DB 192 TLLVVOASGRVTVSTRSQOTIIPNIGSRPMVGGSGRISTW-----TIVKPGD 241
 QY 153 AADIENFASRYLYMAYLYKYTYTNVDEFGASFENKLSFTTGLFGWGIRKALKQIIRSNLP 212
 DB 242 VLIVNSNGN--LIAPRGYKMT-----GKS-----SIMSDAP 273
 QY 213 LDIGIEHSYRLQIHITSS--YKDYMDQIPLPKFAKRSIMV-----VORLLA 259
 DB 274 IDTCISECTTPNGSIPNDKPFQNVNKKITYGACPKYVKQNTLKLATGMRNVPEKQTRGLFG 333
 QY 260 TVAGYVDFPW---YKKWY-----MKLK-----NPMVNRVFPTKKFFN 294
 DB 334 AIAGTIENGMEGIDGWIQGFRIHONSEGTQAADLKSTOAIDQINGKLKRVIEKTEKNEKH 393
 QY 295 ---KEIREPSKALK--EKVSTDTKDLFENKIGQGVDFENKEIRDPKALKKEKVSNDADK 349
 DB 394 QKEKEFSEVEGRIDLEKYEVEDTK-----IDLMS-----YNDVLY 429
 QY 350 LFNKIGQGVDFINNEIRDPKALKIRKYSTGAEDLFENKIGQGVDFINNEIRDPKALK 409
 DB 430 ALEN---QHTIDLTSENMKLEKTKROLRENAEDM-----GNGCF----- 467
 QY 410 IRKYVTEADDFENKIGQGVDF 431
 DB 468 --KIYHKCDNACVESIRNGTYD 487

RESULT 34
 HEMA_IADH2
 ID HEMA_IADH2 STANDARD: PRT: 550 AA.
 AC P12583; 084011;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAI chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/8/80).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC Influenza A virus.
 OK NCBI_TaxID=11358;

[1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=87265458; PubMed=2440178;
 RX kids H., Kawoka Y., Naeve C.W., Webster R.G.;
 RA "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks.";
 RT virology 159:109-119(1987).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HAI AND HAZ) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M16738; AAA43144.1; -
 DR PIR: A27813; HMI180.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR Prodom: PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 8 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61659 MW; A10703ACC9CC353 CRC64;

Query Match 4.4%; Score 104; DB 1; Length 550;
 Best Local Similarity 18.9%; Pred. No. 8;
 Matches 92; Conservative 63; Mismatches 169; Indels 164; Gaps 22;

19 ESDSAAANAYMINSMDYLS-----AVSDNFAERICSOVPGKS 57
 DB ESNMNFNCYPR--DVPDVAISRLVASSGTLFTEGTTWGTGONGSSNACKRGP--- 143
 QY 58 NCSASVSAVMSRCANQDCITLQSLKYPLEAKYQPLTLPDPOLEAFTLFKESDANPANS 117
 DB 144 -----ASGFFSRLL--NMLTKSGSTYPV---LNVTPMNDNFDKLYI----- 179
 QY 118 TEKRFMRFRGKNHSEFHDLYVFLLEKN-----VTRDADATDIENFASRYLYNA--- 167
 DB 180 -----W-----GVHHPSTQEQBOTNLVVOASGRVTVSTRSQOTIIPNIGSRPMVGGSGR 229
 QY 168 -TLVYKYTYTNVDEFGASFENKLSFTTGLFGWGIRKALKQIIRSNLPDIDGIEHSYRLQH 226
 DB 230 ISIIYTWIYKPGDVLIVNSNGNLIAPRGYF--KMTGKSSIMSDAPIDCVSECTTPNGS 287
 QY 227 ITSS--YKDYMDQIPLPKFAKRSIMV-----VORLATIYGVYDTPW---Y 270
 DB 288 IPNDKPFQNVNKKITYGACPKYVKQNTLKLATGMRNVPEKQAGALGALGFIENGMEGT 347
 QY 271 KKWY-----MKLK-----NPMVNRVFPTKKFFN---KEIREPSKALK 305
 DB 348 DQWYGFRIHONSEGTQAADLKSTOAIDQINGKLKRVIEKTEKHQIEKESVEYGRQ 407
 QY 306 --EKVSTDTKDLFENKIGQGVDFENKEIRDPKALKKEKVSNDADKFNKIGQGVDFI 363
 DB 408 DLEKYEVEDTK-----IDLMS-----YNDVLYALEN---QHTIDLT 440
 QY 364 NNEIRDPKALKIRKYSTGAEDLFENKIGQGVDFINNEIRDPKALKIRKYVTEADDFEN 423


```

CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51396: CAA35761.1; -.
DR PIR: S07349; GRKSP1.
DR MGD: MGI:1306778; Mcap1b.
DR InterPro: IPR000102; MAP1B_neuraxin.
DR Pfam: PF00414; MAP1B_neuraxin; 10.
DR PROSITE: PS00230; MAP1B_NEURAXIN; 7.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1874 1890 MAP1B 1.
FT REPEAT 1891 1907 MAP1B 2.
FT REPEAT 1908 1924 MAP1B 3.
FT REPEAT 1925 1941 MAP1B 4.
FT REPEAT 1942 1958 MAP1B 5.
FT REPEAT 1959 1975 MAP1B 6.
FT REPEAT 1993 2009 MAP1B 7.
FT REPEAT 2010 2026 MAP1B 8.
FT REPEAT 2027 2043 MAP1B 9.
FT REPEAT 2044 2060 MAP1B 10.
FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
KKEE AND KKEIV/ REPEATS).
SQ SEQUENCE 2464 AA; 270408 MW; FBD0DD99CFBDBD87 CRC64;
Query Match 4.4%; Score 103.5; DB 1; Length 2464;
Best Local Similarity 20.3%; Pred. NO. 55;
Matches 99; Conservative 77; Mismatches 180; Indels 131; Gaps 21.
OY 15 LAASSVDSAAANVYINIDMSDYLSAVSDNFMERCSQVPRKSNCSASVSAWMSACAD 74
DB 331 LEEERSQGSTSISDMKKNLISPDGLVFLNVPENLKDEP-----NIMKRSIEEA 381
OY 75 CLTLOSLLKYPPEAKYQPTLPDPDYLAEAFILFKESDA-----NPANST-EKRFWR 125
DB 382 CFTLLQYLN-KLSMKREPLFRSVGNITIEPY-ILFGKMGVQKLEMYLVNLPYKSKEMQYEWQ 439
OY 126 FRGRKNHSYFHDVLVNLLEKAVTROADATDIENFASRLYLMATLYKTYITTVNDEGASFF 185
DB 440 QWTGYNKKDAELLILRNGQEVDIR-----I 463
OY 186 NKLSTFTGTFGCGIRAKLQIIRSNMLPDIGTEHSV---SRLOHTTSYKDYMDQIPA 241
DB 464 SYLTGVSLLIVHNPANPAEKIIRVLPFGH-STQYINILEGLKHL-----DPL----- 511
OY 242 LPRFAKRPESLMVQRLAT---VAGYVDTPMYKKWYMKLKNMNVNRFI--PTKKFENKE 296
DB 512 -----KQPLATQKDLTGQVTPRPYKQ--VKLQRADRSRESLPATKPVAKSKS 556
OY 297 IREPK-----ALKER--STPTKDLFEKKIOGTYVDFNFKELRDS 336
DB 557 VAKSEKELTPRYTKTSQVEKTPPKVSKEEVLVKKRPAKTESKPSVTEKEVSKESQSFV 616
OY 337 KA-LKEKVSNDAKDLFENKIGOGTYVDFINNEIR-----DPSKALIRKYSTGAEDLF 386
DB 617 KAEVALEKQATESKP---KVTQKV--VKKEIKTLLEKKKEEKPRKVEYKKKDKTLPKKD 670
OY 387 EKKIIOGYVDFINNEIRDPSSKALIRKVTTEADLEENK--IQOGTYVDFINKEIRDPSSKAL 444
DB 671 EKPRKEEVKKEIKKEIKERKEKLEKVEYKKEETPLDKARKVEYKKEKKEVKKKEKPEKKE- 729
OY 445 IRKVS TE 451
DB 730 IKKISKD 736

```

```

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
GN HA.
OS Influenza A virus (strain A/Duck/Memphis/928/74).
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11367;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; Pubmed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoaka Y., Naeye C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts";
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73772; -; NOT_ANNOTATED_CDS.
DR HSSP; P03437; ZHMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KM Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 566 AA; 63572 MW; C65DFC5DE5BED80D CRC64;

Query Match 4.4%; Score 103; DB 1; Length 566;
Best Local Similarity 17.9%; Pred. No. 9.6;
Matches 92; Conservative 67; Mismatches 166; Indels 188; Gaps 22.

OY 5 DSVGVDTKTLAASESVSAANAAYINSDMSDYLS-----AVSD 43
D 93 DVQDETMDLYERSARSNCYPR---DVDPYASLRSLVASSGTFETEGFTWGTGTV 148
OY 44 NFAETIGQVPRKGSNCSSVSAIYMRCAKQDCITLQSLKYPEAKYQPTLPDPQOLENA 103
D 149 NGGSNAKCRGP-----ASGFESRL--MWLTSGSTYPV---LNVTMENDNFKL 193
OY 104 FI-----LFKESDPAANPSTEK-----RFMRRFRGKNHSYFHDL 138
D 194 YWGVNHNHSTQEDQDLYVQASGRVTVSTRSQQTIIINIGSRPVPVWQSGHLSITW--- 250
OY 139 VFNLEKNTVDADATDIENFASRYLYMATLYKYKTYTVNDEFGASFENKLSFTTGLFGWG 198
D 251 -----TIVRPGVLYVINSNGN--LIAPRGYEF----- 275
OY 199 IKRALKQIIRSNLPDDIGTEHSVSLQHTITS--YKQWMDQIPLALPKFAKFSLMV--- 253
D 276 IRTGKSSIMRSDAPIDTFCISPECTIPNGSIPNDKRPQVNNKLTYYGACQPYVVKOSTLKATG 335

```

```

OY 254 -----VORLLATVAGYVDFPM---YKKWY-----MKLK-----NF 280
DB 336 MRVPEKRTKRGLEFVAIAGTENGEGMIDCWYGFHRONSEGTOAADLSTQAIQING 395
OY 281 MVRVPEIPPKFFENKEIREPSKALKEKYSTDTDFENKIGQGVDF--FNKEIRPSKA 338
DB 396 KLNVRVIEKTEKNEKHOLEKEES-----EVEGRIGDL-ENVVEDTKIDLMSYNAEL---LVA 446
OY 339 LKEKVSNDAKDLFENKIGQGVDFINNEIRDPKSKALIRKYSTGAEDLFENKIGQGVDFI 398
DB 447 LE-----NQHITDITDSEMNKLFEERTRROLRENAEDM-----GSGCF--- 483
OY 399 NNEIRDPKSKALIRKYTEADDLFENKIGQGVDF 431
DB 484 -----KIYHKCDNACIESIRNGTID 503

RESULT 38
MYSH ACACA STANDARD: PRT: 1577 AA.
AC P47808;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE High molecular weight form of myosin I (HMMMI).
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota, Acanthamoebidae, Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91056121; PubMed=2243110;
RA Horowitz J.A., Hammer J.A. III;
RT "A new Acanthamoeba myosin heavy chain. Cloning of the gene and
RL J. Biol. Chem. 265:20646-20652(1990).
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M60954; AAA27709.1; -.
CC DR HSSP: P08799; IMND.
CC DR InterPro: IPR000048; IO.
CC DR InterPro: IPR000857; MYTH4.
CC DR InterPro: IPR001452; SH3.
CC DR InterPro: IPR001609; myosin_head.
CC DR Pfam: PF00612; IO; 1.
CC DR Pfam: PF00063; myosin_head; 2.
CC DR Pfam: PF00784; MYTH4; 1.
CC DR Pfam: PF00018; SH3; 1.
CC DR PRINTS: PRO0193; MYOSINHEAVY.
CC DR ProDom: PD000355; myosin_head; 1.
CC DR SMART: SM00015; IO; 1.
CC DR SMART: SM00242; MYSC; 1.
CC DR SMART: SM00139; MYTH4; 1.
CC DR SMART: SM00326; SH3; 1.
CC DR PROSITE: PS50002; SH3; 1.
CC DR PROSITE: PS50096; IO; 1.
CC DR Myosin: Actin-binding; ATP-binding; Methylation; Multigene family;
CC SH3 domain.
CC KW SH3 domain.
FT DOMAIN 1 754 MYOSIN HEAD-LIKE.
FT DOMAIN 755 782 IO.
FT DOMAIN 1519 1577 SH3.
FT NP_BIND 168 175 ATP (POTENTIAL).
FT DOMAIN 628 650 ACTIN-BINDING (POTENTIAL).

```

```

FT DOMAIN 730 744 ACTIN-BINDING (POTENTIAL).
FT DOMAIN 777 780 POLY-GLY.
FT DOMAIN 1471 1476 POLY-ALA.
FT DOMAIN 1482 1489 POLY-GLN.
FT DOMAIN 1493 1496 POLY-GLN.
FT DOMAIN 1506 1511 POLY-GLN.
FT MOD_RES 119 119 METHYLATION (TRI-) (POTENTIAL).
SQ SEQUENCE 1577 AA; 177529 MW; BAF288EBA8B7438C CRC64;

Query Match 4.4%; Score 103; DB 1; Length 1577;
Best Local Similarity 20.0%; Pred. No. 34;
Matches 98; Conservative 78; Mismatches 185; Indels 130; Gaps 21;

OY 6 SVGDVTKTLTAAESVDSANAMAIMNSMDYLSAYSDNFAEICQVPRGSCMSASYA 65
DB 191 STNGVEONILLANPILFESFNMAKTLRNSSRFKWEIHPDQ-----KGSICGASTIN 243
OY 66 YMSRCAKQDCITLQSLKYPLEAKYOPTLPPDYQLEAFTLFESDANPANSTERKFMNR 125
DB 244 HL-----LEKSRVYQIKGERN-----PRIVATELY---KAP-----R 274
OY 126 FRGKNHSYFHDIVFNILKXNVTRODAT--DIENPASRYLYMATLYKTYTNVDFGASF 184
DB 275 SRGGGSSPARPESFRLSQSCIDVEGVDFKFEFEERYLCHQARVVOFSEDDIN--- 331
OY 185 FNKLSTTGIFGWC-----IKRALKQIIRSLPLDIGT--EHSVSRID 225
DB 332 -NCMELISATLHIGNEFVSGGKNVETSVANREEVKYATLTKVDPALTEQNVTSKLM 390
OY 226 HTTSYKYDMDQIPALPKFRFSIMVVOURLATVAGYVDFPMYKMYKLNFWNRY 285
DB 391 EI-----KGDPTRIPLP-----VQATATNA-----LAKAIYSKLDWLVKRI 430
OY 286 FIFTKFFPKNEIREPSKALKEKYSTDTKDLF-----ENKIGQGVDFEFKKEIRDP----- 335
DB 431 -----NESMEPQKGAR--TTIGVLDFGEIIPDKNSFEDLCINFNEKIQOHFNQY 480
OY 336 SKALKEKVSNDADLFENKIGQGVDFINNEIRDPKSKALIRKYSTGAEDLFENKIGQGV 395
DB 481 TFLLEKLVQSEVKEYEH-----ITFIDN-----PYLDLIEKKQPGGLM 530
OY 396 DFINNEIRPSKA---LIRKYTEADDLFENKIGQGVDFINKE-----IRDPKALI 445
DB 521 LVVDEQSIKSSDANFEIKANQTAARSTQLNGGEDSRPDIILKHVAGVIVDSGMLE 580
OY 446 RKVSTEDNLL 456
DB 581 KNRDTLQKDL 591

RESULT 39
TOP2_SCHPO STANDARD: PRT: 1485 AA.
AC P08096; O74336;
DT 01-AUG-1988 (Rel. 08, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2 OR SPBC1A.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87053875; PubMed=3023070;
RA Oemura T., Morikawa K., Yanagida M.;
RT "The nucleotide sequence of the fission yeast DNA topoisomerase II
RT gene: structural and functional relationships to other DNA
RL topoisomerases."
RL EMBO J. 5:2355-2361(1986).
RN [2]

```


RX MEDLINE-20085745; PubMed-10619025;
 RA Hohenester E., Tisi D., Tails J.F., Timpl R.;
 RT "The crystal structure of a laminin G-like module reveals the
 RT molecular basis of alpha-dystroglycan binding to laminins, perlecan,
 RT and agrin.";
 RL Mol. Cell 4:783-792(1999).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-
 CC 4 (S-MEROSIN).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MORINE MUSCULAR DYSTROPHY
 CC (DY2J).
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: U12147; AAC52165.1; -;
 DR EMBL: X69869; CAA49502.1; -;
 DR EMBL: S75315; AAB33573.1; -;
 DR PDB: 1000; 03-DEC-99.
 DR MGD: MGI:99912; Lama2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001886; LAMNT.
 DR InterPro: IPR000034; Laminin_B.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam: PF00052; Laminin_B; 2.
 DR Pfam: PF00053; Laminin_EGF; 15.
 DR Pfam: PF00054; Laminin_G; 5.
 DR Pfam: PF00055; Laminin_Nterm; 1.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR ProDom: PD002082; LAMNT; 1.
 DR ProDom: PD003031; Laminin_B; 2.
 DR SMART: SM00180; EGF_Lam; 15.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00281; Lamb; 2.
 DR SMART: SM00282; Lamb; 5.
 DR SMART: SM00136; LAMNT; 1.
 DR PROSITE: PS00022; EGF_1; 11.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01248; LAMININ TYPE EGF; 14.
 DR PROSITE: PS00025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 22
 FT CHAIN 23 3106 LAMININ ALPHA-2 CHAIN.
 FT DOMAIN 23 282 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
 FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
 FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
 FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
 FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).

FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 752 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
 FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
 FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
 FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
 FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
 FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.
 FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
 FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
 FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14 (N-TERMINAL).
 FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1376 1415 LAMININ EGF-LIKE 1.
 FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15 (C-TERMINAL).
 FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
 FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
 FT DOMAIN 1570 2140 LAMININ II AND I.
 FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
 FT DOMAIN 2326 2517 LAMININ G-LIKE 2.
 FT DOMAIN 2522 2706 LAMININ G-LIKE 3.
 FT DOMAIN 2759 2930 LAMININ G-LIKE 4.
 FT DOMAIN 2929 3106 LAMININ G-LIKE 5.
 FT DOMAIN 1662 1863 LAMININ G-LIKE 6.
 FT DOMAIN 1923 2146 COILED COIL (POTENTIAL).
 FT DOMAIN 283 292 COILED COIL (POTENTIAL).
 FT DISULFID 285 303 BY SIMILARITY.
 FT DISULFID 305 314 BY SIMILARITY.
 FT DISULFID 317 337 BY SIMILARITY.
 FT DISULFID 340 349 BY SIMILARITY.
 FT DISULFID 342 374 BY SIMILARITY.
 FT DISULFID 377 386 BY SIMILARITY.
 FT DISULFID 389 407 BY SIMILARITY.
 FT DISULFID 410 422 BY SIMILARITY.
 FT DISULFID 412 438 BY SIMILARITY.
 FT DISULFID 440 449 BY SIMILARITY.
 FT DISULFID 452 462 BY SIMILARITY.
 FT DISULFID 465 478 BY SIMILARITY.
 FT DISULFID 467 482 BY SIMILARITY.
 FT DISULFID 484 493 BY SIMILARITY.
 FT DISULFID 496 511 BY SIMILARITY.
 FT DISULFID 511 511 BY SIMILARITY.
 FT DISULFID 511 762 BY SIMILARITY.
 FT DISULFID 755 769 BY SIMILARITY.
 FT DISULFID 772 781 BY SIMILARITY.
 FT DISULFID 784 800 BY SIMILARITY.
 FT DISULFID 803 838 BY SIMILARITY.
 FT DISULFID 805 828 BY SIMILARITY.
 FT DISULFID 831 840 BY SIMILARITY.
 FT DISULFID 843 858 BY SIMILARITY.
 FT DISULFID 861 875 BY SIMILARITY.
 FT DISULFID 863 882 BY SIMILARITY.
 FT DISULFID 885 894 BY SIMILARITY.
 FT DISULFID 897 911 BY SIMILARITY.
 FT DISULFID 914 926 BY SIMILARITY.
 FT DISULFID 916 933 BY SIMILARITY.
 FT DISULFID 935 944 BY SIMILARITY.
 FT DISULFID 947 960 BY SIMILARITY.
 FT DISULFID 963 975 BY SIMILARITY.
 FT DISULFID 965 981 BY SIMILARITY.
 FT DISULFID 983 992 BY SIMILARITY.
 FT DISULFID 995 1007 BY SIMILARITY.
 FT DISULFID 1010 1019 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1028 1037 BY SIMILARITY.
 FT DISULFID 1040 1053 BY SIMILARITY.
 FT DISULFID 1056 1068 BY SIMILARITY.
 FT DISULFID 1058 1075 BY SIMILARITY.
 FT DISULFID 1077 1089 BY SIMILARITY.
 FT DISULFID 1089 1099 BY SIMILARITY.
 FT DISULFID 1416 1425 BY SIMILARITY.
 FT DISULFID 1418 1432 BY SIMILARITY.
 FT DISULFID 1435 1444 BY SIMILARITY.
 FT DISULFID 1447 1462 BY SIMILARITY.
 FT DISULFID 1465 1480 BY SIMILARITY.

```

FT DISULFID 1467 1490 BY SIMILARITY.
FT DISULFID 1493 1502 BY SIMILARITY.
FT DISULFID 1505 1520 BY SIMILARITY.
FT DISULFID 1523 1535 BY SIMILARITY.
FT DISULFID 1525 1542 BY SIMILARITY.
FT DISULFID 1544 1553 BY SIMILARITY.
FT DISULFID 1556 1567 BY SIMILARITY.
FT DISULFID 1570 1570 INTERCHAIN (PROBABLE).
FT DISULFID 1574 1574 INTERCHAIN (PROBABLE).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match 4.38; Score 102.5; DB 1; Length 3106;

Best Local Similarity 20.3%; Pred. No. 85;

Matches 109; Conservative 68; Mismatches 178; Indels 183; Gaps 26;

```

QY 4 SDSYGVYTKTLASSESANANWIMSDYLSVSDFAIRICQVPGKSGCSASY 63
Db 1671 AESLEEFIKGLVQDAEINE--KAVKLNETIGN-----QDKTAERNLELOK-----EI 1717
QY 64 SAYMSRCAKODCLTLQSLKYFLAKYQPLTLPDYOLEAFILFKESDANPANSTERKRW 123
Db 1718 DRMLKELRSKDLQT-----QKEVA-----EDELVAAGELKKRVN 1751
QY 124 MRF--RGKKNHSYFDLVFNLE-KNVTROA-----DATDIENFASRYLYMATLYKTYT 175
Db 1752 KLFGEPRQONDEMEKDLQOKLAEYKNNKLDAMDLLRENTDTRDANR--LSAANQKNMT 1808
QY 176 NVDEFGASFEFNKLSFTTGLFGMGIRALKQIIR-SNLPDIGHSEHSRLQHTSSYKDY 234
Db 1809 ILETKKKEAI-----EGSKROIENTLKEGNDILD-----EANOGLGEINSVIDY 1851
QY 235 MDTOIPALPKFAKRFPSLMV-----VQRL-----ATVAGYVDTFW 269
Db 1852 VDDIKTKLPMSSELSDKIDDLAQEIKDRLEKVFQASHAAQLNDSAVLDGILD--- 1908
QY 270 YKKWYMKLKNMNV-----RVFIPTKKFENKEIRPSKALKEKYSTDJKDLFENKIGGT 324
Db 1909 -----EAKNISFNATAAFRAYSNIKDIIDE-----AEKVAREAKEL-----AOGA 1948
QY 325 VDFNKEIRDPKALKE-----KVSNDAK---DLFENKIGGTVDFI 363
Db 1949 -----TKLATSPQGLLKEKDAKSLQKSPRILNEAKKLANDVKGHNHNDLNDLTKRLETAIDL 2004
QY 364 N-----NEIRDPKALIRKYSTGAEDLFENKIGGTVDFINNEIRDPKALIRKYTE 416
Db 2005 NSGLGALNDTMDKLSAITNDTAAKLAQAVKEKA-----REANDTAKAVLAQV--- 2051
QY 417 ADDELLENKIG-----QGTVDFINKEIRDPK-----ALIRKYSTEADNLEK 458
Db 2052 -KDLHQNLQGLKQNYKNKLADSVAKTNNAVVPKSPKIIADAGTSYVNLQEDADRLLDK 2108

```

Search completed: October 19, 2002, 07:49:52

Job time : 36 secs